PERIODICAL
GENERAL LIBRARY
UNIV. OF MICH.
FEB 24 1930

The Journal of EDUCATIONAL PSYCHOLOGY

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching.

BOARD OF EDITORS
HAROLD RUGG, *Chairman.*

J. CARLETON BELL — ARTHUR I. GATES — RUDOLF PINTNER — L. M. TERMAN
FRANK N. FREEMAN — V. A. C. HENMON — BEARDSLEY RUML — E. L. THORNDIKE

VOL. XX
No. 9

DECEMBER, 1929

$5.00 a year.
70c per copy.

CONTENTS

Index and Title-page to Volume XX (1929) i
BIRTH ORDER AND INTELLIGENCE. *L. L. Thurstone and Richard L. Jenkins*. 641
AN INDEX OF INTEGRATION. *John W. Dickey*. 652
SOME CONCLUSIONS FROM A COMPARISON OF SCHOOL ACHIEVEMENT OF CERTAIN RACIAL GROUPS. *Helen G. Pratt* 661
A NOTE ON THE CORRECTNESS OF CERTAIN ERROR FORMULAS. *Karl J. Holzinger* 669
INCIDENTAL LEARNING. *Raymond R. Willoughby* 671
FACTORS INFLUENCING RETENTION OF HISTORY IN THE SIXTH, SEVENTH AND EIGHTH GRADES. *Sarah Janet Bassett*. . . . 683
NOTE ON RELIABILITY PREDICTIONS FOR THE SEASHORE MUSIC TESTS. *Lyle H. Lanier*. 691
REPLY TO DR. LANIER'S NOTE ON THE SEASHORE TESTS. *Paul R. Farnsworth* 693
PHOTOGRAPHIC RECORDS OF EYE-MOVEMENTS IN STUDYING SPELLING. *Ethel M. Abernethy*. 695
PREDICTING ABILITY IN ART. *Alfred S. Lewerenz* 702
NOTE ON THE STANDARD ERROR OF A RELIABILITY COEFFICIENT ESTIMATED FROM THE COEFFICIENT FOR A DIFFERENT RANGE OF TALENT. *Jack W. Dunlap and Edward E. Cureton* . . . 705
NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES. 707
NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION 717

Published Monthly Except June to August by

Warwick & York, Inc.,

10 East Centre Street — Baltimore, Md.

Entered as Second Class matter November 15, 1921, at the Post Office at Baltimore, Md., under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

The Journal of Educational Psychology

is published monthly, except June, July, and August by Warwick & York, Inc., Baltimore, Md., and York, Pa. Title page and index are bound in the December number.

The price of the Journal is $4.00 a year, payable in advance. Foreign postage is 30 cents extra. Single current issues are 60 cents each, and less than a full year's subscription is charged at single issue rate for every month ordered. Back volumes $4.50 each; single back issues 75 cents.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which such change is to take effect. No claim for non-receipt of an issue will be entertained unless made within two weeks after receipt of the next succeeding issue.

Manuscripts for publication may be submitted to any one of the Editors or to the publishers direct. Books for review should be addressed to the publishers, 10 E. Centre Street, Baltimore, Md.

Warwick & York, Inc.,
Baltimore, Md.

TEACHERS' HEADQUARTERS

European Plan

Entirely Fire-Proof

Centrally Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

Thousands of Successful Teachers use

WEBSTER'S NEW INTERNATIONAL DICTIONARY

"The Supreme Authority"

To get accurate, encyclopedic, up-to-date information of all kinds that is of vital use and interest in the schoolroom.

A wise school superintendent has said: *"I have never yet seen a person, whether pupil or teacher, who was accustomed to the frequent use of the dictionary who was not at the same time a good or superior all-round scholar."* A better test than this of the value of dictionary work could not be found.

The New International is constantly revised and improved to keep abreast of modern needs and information. 452,000 Entries, including thousands of new words. 6000 Illustrations. 2,700 Pages.

Write for Helps in Teaching the Dictionary, FREE

G. & C. MERRIAM CO.
Springfield, Mass.

SCHOOL AND SOCIETY

A weekly journal covering the field of education in relation to the problems of American democracy.

The objects of SCHOOL AND SOCIETY are the advancement of education as a science and the adjustment of our lower and higher schools to the needs of modern life.

It encourages discussion of all problems from every point of view; it pays special attention to foreign affairs; it publishes accounts of educational events of the past week and research on fundamental problems.

Subscription $5.00—Single Numbers 15 Cents

[*SCHOOL AND SOCIETY will be sent to new subscribers for thirteen weeks for one dollar, which may be sent at the risk of the publisher.*]

THE SCIENCE PRESS
Grand Central Terminal - - New York, N. Y.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XX December, 1929 Number 9

BIRTH ORDER AND INTELLIGENCE[1]

L. L. THURSTONE

University of Chicago,

AND

RICHARD L. JENKINS

Institute for Juvenile Research

The records at the Institute for Juvenile Research seem to be admirably suited for the present investigation. There are on file more than ten thousand case records which have been accumulated during the last ten years on children varying in age from one year to twenty-one years. One of the routine requirements for practically every case study is a Stanford-Binet examination which is only occasionally varied by introducing some other revision of the Binet examination. All of the mental ages and intelligence quotients that have been used in this study are based on individual psychological examinations. No group tests have been used for this study.

From the case history it is relatively easy to ascertain the number of siblings for each patient and their ages. The cross references made it possible also to pair the siblings for those families which have sent more than one child to the institute for examination. In assigning ordinal number of birth to each child, the question naturally arose whether miscarriages should be counted. In many cases the records are rather complete in this regard but in many others the information is of doubtful accuracy. It was finally decided that the miscarriages

[1] This project was started as one of a series of studies carried out under the auspices of the Illinois Institute for Juvenile Research, Series B, No. 140. It was completed as a project of the Local Community Research Committee at the University of Chicago.

would not be taken into consideration in assigning birth number to each patient because of the fact that all of the case records were not equally trustworthy on this point. The ordinal number assigned to each child takes into consideration all of the siblings who are listed in the case history including deaths but not miscarriages.

The first step in the investigation was to draw a sample of about 1500 cases in chronological order from the records and to tabulate the gross data without any corrections or statistical refinements.

TABLE I.—DISTRIBUTION OF IQ FOR EACH ORDINAL NUMBER

Ordinal number	0–9	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–99	100–109	110–119	120–129	130–139	*n*	*m*	*σ*
1	3	6	11	11	22	38	81	100	87	73	40	21	5	2	500	76.900	22.4
2	1	6	6	9	12	19	58	51	62	48	18	7	3	3	303	76.056	22.3
3	1	1	7	6	9	19	40	46	31	27	20	6	2	..	215	75.093	21.8
4	1	4	6	5	4	20	26	26	24	17	11	6	2	..	152	72.763	24.3
5	..	2	5	1	7	6	9	19	18	8	2	3	..	..	80	71.125	22.7
6	..	1	..	2	2	7	13	5	5	5	1	..	..	1	42	69.524	21.0
7	..	..	2	..	1	2	2	6	5	2	1	..	1	..	22	74.091	23.0
8	..	1	..	3	1	1	2	2	3	2	2	..	..	..	17	67.941	26.5
9	1	..	..	..	..	1	1	1	2	4	..	1	..	..	11	78.636	28.1
10	..	..	..	..	2	1	1	..	..	2	..	..	..	..	6		
11	..	..	..	..	..	1	..	..	..	..	..	..	..	..	1		
12	..	..	..	..	..	..	1	..	..	1	..	..	..	..	2		
13 and over	..	..	1	..	..	..	..	1	1	..	..	..	..	..	3		
Unknown	1	..	..	..	3	6	11	25	11	13	3	2	1	..	76	77.632	18.2
Total	8	21	38	37	63	121	245	282	249	202	98	46	14	6	1430		

This has been reproduced in Table I. The scale of intelligence quotients ranges from 0 up to 140 in class intervals of ten. The ordinal number of the patient was listed up to twelve with two additional categories, namely "thirteen and over" and "unknown." In this table are also listed the number of cases for each birth number, the mean intelligence quotient for each birth number, and the standard deviation of the intelligence quotients for each birth number.

Inspection of this table shows immediately that the number of cases decreases markedly with increase in size of family and this is, of course, to be expected. The calculation of mean intelligence quotient and its standard deviation are not reproduced for the categories with less than ten cases. Our interest will be directed primarily to those categories for which the class frequencies are at least forty or more.

The column of mean intelligence quotients shows that the mean intelligence of the children decreases with increase in ordinal number or with increase in size of family, depending on the interpretation. This relation is presented graphically in Figure 1 which shows that the mean intelligence quotient decreases as ordinal number of the patient increases. The latter part of Table I shows a number of inversions of the first order in the column of means. These inversions occur only for the small class frequencies which have been ignored in the graph. The graph was extended only through the sixth child

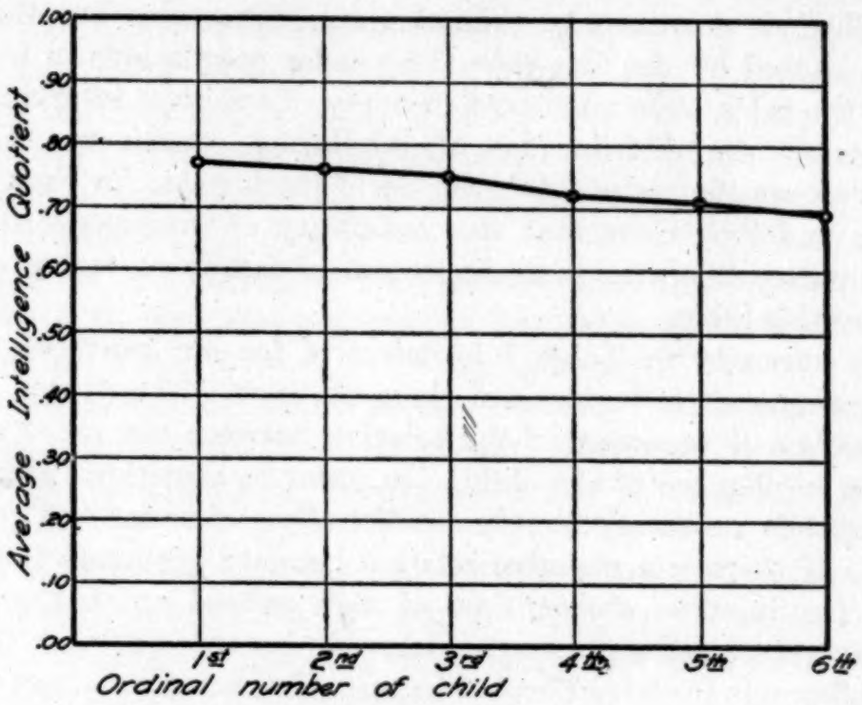


FIG. 1.

where the frequency was forty-two and for the first six ordinal numbers there are no inversions in the means. Beyond that, the frequencies are so small that the mean intelligence quotients are of no interest.

The interpretation of this first gross tabulation is probably that there is a negative correlation between size of family and the mean intelligence of the children. Such a correlation would produce the results that we find here in Table I. It is reasonable to expect that the children of large families in this population, in which the parents are very frequently foreign born, would have slightly lower performance on the Binet tests than the children of smaller families in which the parents are more frequently native born. It is also probably true

that large families represent, on the average, a lower intellectual, social, and economic status than small families. The correlation represented by Table I is, in fact, −.086.

Any correlation between order of birth and intelligence would of course also affect the results of Figure 1. It is clear that before any conclusions can be reached on the present problem of order of birth, it is necessary to control not only the size of the family but also the social and economic status of the parents. This has been done and the results will be described in subsequent sections of this report.

The first glance at Table I gives the impression that the variability of intelligence decreases as ordinal number increases but this is an illusion caused by the fact that the smaller populations in the lower half of the table have no entries in many of the class intervals in the extremes of the distributions of intelligence quotients. The last column shows the standard deviation of intelligence for each ordinal number and it is clear that the variability of intelligence does not change noticeably with increase in size of family as far as one can judge by this table.

The summary in Table I is defective for our purposes in that several antagonistic factors are there at work without any control. Our problem is to ascertain the relation between the order of birth and the intelligence of the child. In order to ascertain this relation it is certainly necessary to rule out the effect of social and economic status. If there is a negative relation between economic status and size of family, then the children of high ordinal number will be of lower average intelligence, not necessarily because they are the later born children in their families but because they are drawn from families of lower social status.

In order to eliminate this disturbing factor we have limited our study to siblings. For example, we have records from 382 families whose first-born and second-born children have been examined at the Institute. Now if we compare the intelligence of first-born and second-born children in these records we shall have eliminated the effect of social and economic status because every family that is represented in the data for the first-born is, of course, also represented in the data for the second-born children. Our comparisons in this study are based entirely on siblings.

If the intelligence of children is improved by the experience of parents in bringing up children, then it is conceivable that such experience would affect the comparison of the intelligence of first and second

born children. The comparison would, of course, be favorable to the second-born child if this factor has any appreciable effect. It is also conceivable that the economic status of the family rises, on the average, with time and that such a factor may influence the intelligence of the children. It is more probable, however, that the causal relation is more strongly in the reverse direction, namely that the most intelligent parents have, on the average, the best economic status, and that the children are bright primarily because the parents are bright and only secondarily because the parents are economically successful. Our principal task is to ascertain the facts. The results may be susceptible of various interpretations.

Having limited ourselves to the study of siblings for the present problem, our first interest is to ascertain how many pairs of siblings are available in our data. The first line of Table II gives the principal facts for the comparison of first-born and second-born children in the same families. There were three hundred eighty-two families for whom we had records of both the first-born and the second-born children. The first two columns of the table show the two birth orders of each comparison. For example, there were one hundred sixty-five families for whom we had records of both the third-born and the fourth-born siblings. There were 58 families for whom we had records of the intelligence quotients of the second- and fifth-born siblings.

The plan of the analysis may be illustrated by the comparison of the first- and second-born siblings. Referring again to Table II the mean intelligence quotient of the first-born children in the three hundred eighty-two families was 81.75 while the mean intelligence quotient of the second-born children in the same families was 84.84. The comparison shows a slight advantage for the second-born. The social and economic status was very nearly the same in these two groups because of the fact that every child in the first-born group had a second-born sibling in the other group. It seems hardly necessary to consider the size of the family as a disturbing variable because the first- and second-born children were drawn from the same families. This arrangement of the present investigation seems to the authors to be far superior to any statistical attempt to control economic status and size of family when the first-born and second-born children are drawn from different families. Fortunately our data were so extensive that we could afford to limit ourselves to comparisons of siblings only.

TABLE II

Birth orders		n	m_a	m_b	σ_a	σ_b	d_{a-b}
a	b						
1	2	382	81.75	84.84	18.58	19.48	−3.09
1	3	195	78.24	82.76	17.40	21.19	−4.52
1	4	94	75.13	85.04	16.17	17.05	−9.91
1	5	43	74.72	79.37	15.41	20.98	−4.65
1	6	20	72.80	74.10	14.28	24.18	−1.30
1	7	6	71.33	81.33	.....	.....	−10.00
1	L	2	52.00	56.00	.....	.....	−4.00
2	3	253	82.44	83.78	18.21	19.61	−1.34
2	4	130	78.94	85.29	15.99	17.32	−6.35
2	5	58	75.83	83.34	17.30	20.31	−7.51
2	6	21	75.86	75.29	15.33	12.13	+.57
2	7	12	68.83	83.00	12.15	13.14	−14.17
2	L	5	59.00	70.20	.....	.....	−11.20
3	4	165	82.62	84.31	20.81	20.35	−1.69
3	5	72	77.83	82.72	20.48	19.08	−4.89
3	6	24	78.33	76.00	17.05	20.05	+2.33
3	7	12	76.00	73.67	17.50	12.71	+2.33
3	L	4	55.50	73.00	.....	.....	−17.50
4	5	103	81.06	86.48	18.35	18.59	−5.42
4	6	39	72.13	77.62	18.04	21.55	−5.49
4	7	13	71.62	78.08	21.11	19.18	−6.46
4	L	3	62.33	70.33	.....	.....	−8.00
5	6	62	79.52	76.97	18.11	21.76	+2.55
5	7	22	76.18	75.55	16.74	19.97	+.63
5	L	10	73.00	76.80	17.75	22.78	−3.80
6	7	28	74.57	77.86	17.62	19.47	−3.29
6	L	12	77.50	82.50	14.59	20.79	−5.00
7	L	14	77.00	88.00	13.56	16.83	−11.00

The next two columns of Table II show the standard deviations of the intelligence quotients in the two groups, the first-born and the second-born. The variability of intelligence is slightly greater among the second-born than among the first-born. The last column of the table shows the difference between the two mean intelligence quo-

tients. The average difference between the intelligence quotients of the first-born and the second-born is 3.09 points in favor of the second-born siblings. Similar comparisons are made in each of the twenty-eight lines of Table II.

Our comparisons were made up to and including the eighth-born siblings. With eight birth orders it is clear that we have $\frac{n(n-1)}{2} = 28$ possible combinations of birth orders. These are represented in the twenty-eight lines of Table II.

Our principal findings are summarized in the last column of Table II. Here we find that $m_1 - m_2 = -3.09$, $m_1 - m_3 = -4.52$, $m_1 - m_4 = -9.91$, all of which favor the later born sibling. The five exceptions in which the difference is recorded with positive sign represent relatively small groups of families. There were sixty-two families for whom we had records of the fifth- and sixth-born siblings and this comparison favors the fifth-born rather than the sixth-born. The fact that twenty-three out of the twenty-eight comparisons favor the later born siblings and the further fact that all of the comparisons involving large samples consistently favor the last born seem to justify the conclusion that intelligence increases on the average with order of birth in the same family.

Another finding of some interest is that the variability of intelligence seems to increase slightly with birth order. The standard deviations were not calculated for samples as small as ten and they are of no consequence for samples less than forty or fifty families. These comparisons are shown in the sixth and seventh columns. In eighteen of the twenty-three comparisons the variability of intelligence is slightly greater for the later born siblings than for the younger born siblings.

Our next problem is to study the internal consistency of the differences recorded in the last column of Table II. If the difference between the first-born and the second-born is 3.09 points in IQ and if the difference between the second-born and the third-born is 1.34 points in IQ, then the difference between the first-born and the third-born should be the sum of these two differences, namely 4.43 points in IQ. But the direct comparison of these two birth orders, the first and the third, shows a difference of 4.52 points. This is a fair degree of consistency but the consistency is not nearly so good when the comparisons are made for higher birth orders where the samples were small.

Since the samples are fairly large for the comparisons that involve the first four birth orders we shall make a separate analysis of them. We want to use all of the available comparisons in Table II in order to ascertain the best fitting increments in intelligence quotients for the successive birth orders. We shall do this first for the first four birth orders. In Table III we have listed the six observation equations for the first four birth orders. These are obtained directly from Table II.

TABLE III.—OBSERVATION EQUATIONS

	m_1	m_2	m_3	m_4	d	n
12	1	−1	...	...	3.09	382
13	1	...	−1	...	4.52	195
14	1	...	...	−1	9.91	94
23	..	1	−1	...	1.34	253
24	..	1	...	−1	6.35	130
34	..	...	1	−1	1.69	165

The last column of Table III shows the number of families represented in each comparison and this will be used as the weight for each observation equation. In Table III the variables are indicated at the top of the columns so that only the coefficients are listed in the body of the table.

These six observations equations were weighted by the number of families as shown in Table III. They were then solved by the method of least squares in the usual manner. In order to solve them it was necessary to adopt an origin of measurement. The mean intelligence quotient of the first-born was adopted as an arbitrary origin, or datum. The mean intelligence of each of the other birth orders is expressed as an increment in terms of points of IQ measured from the mean intelligence of the first-born as a datum. In solving the equations the value of m_1 is therefore zero. The least square solution is as follows: $m_1 = 0.0000$; $m_2 = 3.1112$; $m_3 = 4.8725$; $m_4 = 8.0459$. These best fitting values show again that the mean intelligence increases with birth order.

Assuming the above values and solving for the best fitting values for the remaining four birth orders by the method of least squares, we obtain the values listed in Table IV. These data have been analyzed by several different statistical methods and they all give practically the same results. In Figure 2 we have a graphical representation of the increase in mean intelligence with birth order.

TABLE IV

m_1	0.0000
m_2	3.1112
m_3	4.8725
m_4	8.0459
m_5	11.2812
m_6	8.4285
m_7	10.8668
m_L	17.4167

The inference seems natural here that on the whole the later born siblings tend to be on the average brighter than the first-born. Not only does this seem to be the case in the comparison of the first-born with the subsequent children, but the rise in intelligence with order of birth seems to continue as far as the eighth-born child. The data seem to indicate that the differences are not caused merely by

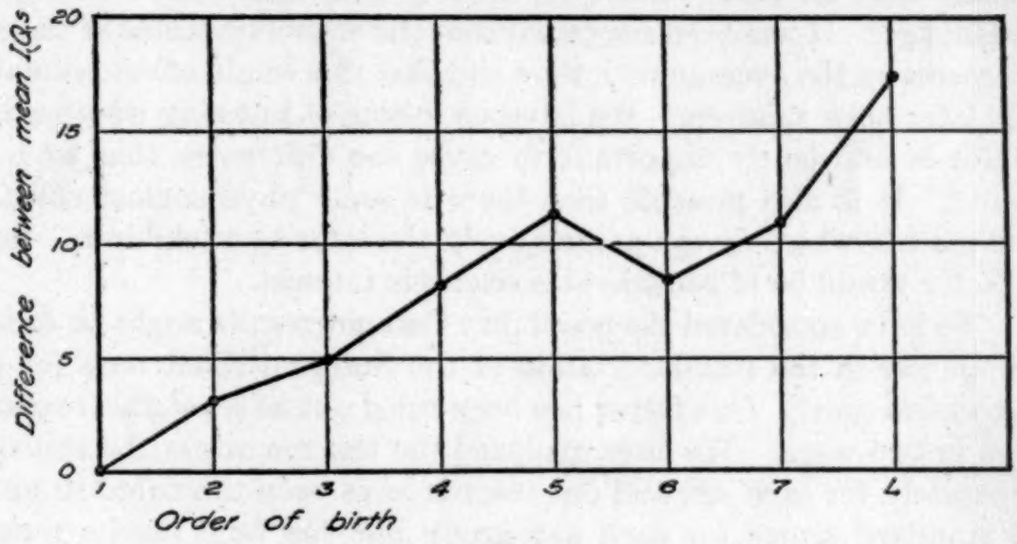


FIG. 2.

some handicapping factor that affects only the first-born child. Whatever may be the explanation, the rise in intelligence seems to continue for the successive children in the same family.

If there is no peculiar selective factor which brings into the institute clinic the bright younger children of a family or which keeps out the dull younger children of the same family, then the present tabulations would support the medical findings which generally indicate that the first-born child in a family is more likely to be handicapped than the later born children. Our results indicate the additional interesting fact, however, that the intelligence quotients of successive children in the same family are not only unfavorable to the first-born child

particularly but that the effect seems to be progressive whereby the mean intelligence rises with order of birth.

It is not at all the purpose of this investigation to attempt an explanation of this interesting fact. Explanations will be found undoubtedly in great number and they will probably include both hereditary and environmental factors. As far as we know there is nothing in our data to show conclusively which of these two general types of explanation is the more plausible. It is possible that experience in bringing up one or more children has a favorable effect on the test intelligence of later born children but it seems strange that such an effect should persist so as to be noticeable in children of school age. It is known that later born children learn to talk on the average at an earlier age than first-born children and while this factor would have an effect on intelligence quotients of very young children it can hardly have an effect sufficiently lasting to be measurable during the school age. It has been suggested that the economic status of parents increases on the average with time and that this would affect favorably the later born children. We have no means of knowing whether this factor is sufficiently important to cause the differences that we have found. It is also possible that there is some physiological effect in the mother which favors progressively the later born children. Such a factor would be of considerable scientific interest.

We have considered the possibility that our results might be caused by defects in the standardization of the Stanford-Binet tests for the successive ages. This factor has been ruled out as a possible explanation in two ways. We have analyzed the test records of the institute separately for each age and our records have been tabulated in terms of standard scores for each age group but the final results remain the same. A further check on our findings is a study in progress by Miss Minnie Steckel in Sioux City, Iowa. Miss Steckel has given four group tests to the public school population in that city and the records of all siblings have been compared. Since that study involved the testing of a large school population, the tests were standardized on the experimental population itself. The published norms were not used. The test performances were reduced to standard scores for each age and the comparisons were limited again to the siblings. The preliminary findings in that study agree with the findings of the present investigation.[1] The institute population is retarded in that

[1] Miss Steckel's investigation will be published as a doctor's thesis at the University of Chicago.

its mean intelligence quotient is about .80. Although the mean intelligence quotient of our population is lower than the average for the general population we have seen no very good reason to suspect that this fact would in any way influence our main problem.

The present investigation seems to justify the conclusion that the mean intelligence quotient increases with birth order and that this effect is not limited to a handicapping of the first-born. The effect seems to be progressive at least as far as the eighth-born child. The later born children seem to be brighter on the average than their earlier born siblings and the variability of test intelligence seems to increase with order of birth.

AN INDEX OF INTEGRATION

JOHN W. DICKEY

Palmerton, Pa.

Aim

The purpose of this paper is to study pupil integration within the public schools by the use of a mathematical formula. In order to do this, it seemed to me to be very desirable to have a valid and reliable formula for measuring the degree of this integration. An attempt was made to set up a formula so that the resulting indexes would be at once comparable. After spending considerable time and energy in an unsuccessful search for such a formula, as a last resort, realizing that it is a somewhat precarious thing to do, I have nevertheless attempted to set up such a formula which will be explained somewhat in detail later on.

The importance of pupil integration has been raised by Bagley when he says that the early introduction of the factor of differentiation may serve to limit the factor of integration providing "a common basis of certain ideas and ideals and standards which go a long way toward insuring 'social solidarity'—*a basis of common thought and common aspiration which is absolutely essential to an effective democracy.*"[1]

The integrating function may also extend itself into secondary education, for as Inglis states, "From this (social cohesion) arises the integrating function of secondary education, which in this country particularly is constantly acquiring greater and greater importance for a number 4 of reasons."[2] The integration concept (like-mindedness) is, for these reasons, somewhat universal.

Definition of Integration

From the point of view of objective psychology, a quite satisfactory concept of integration, which lends itself to statistical treatment, would be as follows: Each course of study, reading, writing, arithmetic, etc., is a network of pattern-responses. Two plus four is six. "Six" is the pattern-response for the situation "two plus four." The word "bell" is a situation which brings to the pupils mind a pattern-response

[1] Bagley, W. C.: The "Six-six" Plan. *School and Home Education*, Vol. XXXIV, p. 3.

[2] Inglis, Alexander: "Principles of Secondary Education." P. 297.

similar in core thought at least to the reader's response, after the pupil-learning has taken place. All systems of knowledge may be placed upon this basis.

The Formula

No single test could include all of the pattern-responses which constitute a given course of study. The securing of a representative sample is one way out of the difficulty. Assuming that the measurement of human traits furnishes data which conform very closely to the theoretical probability curve,[1] we have a criterion to judge our sampling. As will later be shown, that the facts agree to a considerable extent with the theory, we are reasonably safe in concluding that the tests used adequately measured the quantitative mastery of the pattern-responses. The degree of quantitative mastery is, by our definition, the degree of integration.

It is a perfectly normal assumption that there is a need for the measurement of integration. Its uses are somewhat numerous including the comparison of grades within a given school system and the comparison of one school system with another. When comparing school systems, on this basis, the forms of standardized achievement and intelligence tests acquire additional significance. The measurement of integration, calling attention to it, may add to its control.

If the pupils of a given grade had learned all of the pattern-responses of arithmetic given in that grade, the standard deviation would equal zero, and the integration symbol would equal infinity. If, to the contrary, they had learned none of these pattern-responses the integration symbol would equal zero. The range of integration, in arithmetic or in any other subject, would be from zero to infinity.[2]

In the first case (complete integration) the deviations from the mean are zero, making the standard deviation equal to zero.

In the second case (no integration) the deviations from the mean are again zero, also making the standard deviation equal to zero.

The following formula, including the two most important measures of a distribution (central tendency and dispersion), satisfies these conditions ("complete integration" and "no integration").

$$K = \frac{M}{SD}$$

---

[1] Holzinger, Karl J.: "Statistical Methods for Students in Education." P. 204, from Cyril Burt, "Mental and Scholastic Tests," 1921.

[2] In this and the following discussions, it is assumed that the probable errors of the measures involved are present, but no attempt is made to deal with them.

The letter K is the symbol for the index of integration;[1] the letter M, the symbol for the arithmetic mean; and SD, the symbol for the standard deviation of the mean.

Having both limits formulated, it remains to express, functionally, the measures that lie between these limits. It was observed while experimenting with the probability curve, that integration was the reciprocal of dispersion. Our formula is one hundred times the reciprocal of Pearson's coefficient of variation (V) which is an abstract number, and a relative measure of dispersion. Since the reciprocal of an abstract number is also an abstract number, we satisfy this functional relationship. We therefore have a formula which represents the relative degrees of integration and places the indexes on a comparable basis,[2] just as are "four per cent" and "six per cent."

---

[1] The letter K was chosen since its broken line closely resembles a line graph of the indexes of integration for various ages, from the first grade through high school to the university, as was found from examination of some data.

This is called an index because it is a reversible ratio involving the arithmetic average, and because the results secured parallel general thinking tendencies. (See Kelley: "Statistical Method." P. 334.)

[2] This somewhat technical proof is given to show that the values gotton from the formula are directly comparable.

Both M and SD are independent variables since M is a function of the sum of the scores and the population; and SD is a function of the deviations about the mean and the population.

Securing the partial derivatives of

$$K = \frac{M}{SD}$$

we get these two equations:

$$\frac{\partial K}{\partial M} = (SD)^{-1}$$

treating SD as a constant; and

$$\frac{\partial K}{\partial (SD)} = -M(SD)^{-2}$$

treating M as a constant.

Interpreting each of these as a ratio of time rate of change, first of K to that of M, and second of K to that of SD, we are consistent when the ratio of these two ratios, namely,

$$\frac{\partial K/\partial (SD)}{\partial K/\partial M} = \frac{\partial M}{\partial (SD)} = -\frac{M}{SD}$$

is interpreted as the ratio of the time rate of change of M to that of SD. The minus sign can be neglected since it designates direction. Our original equation determines values which are directly comparable, just as rates are comparable.

By partial integration of the formula, and substitution of values secured from "complete integration" and "no integration," the earlier stated conclusions are immediately seen to be valid for these two limits.

In using this formula to compare norms, the measures of central tendency may be expressed in terms of the mean, the median or the fifty-percentile, and less frequently the mode. In a theoretical distribution they are identical measures and are interchangeable. In actuality, they may likewise be substituted when the population is sufficiently large to present negligible skewness.

As substitution for the measures of dispersion, the following relationships would be adequately satisfactory:

$$SD = 1.4826,\ Q = 1.4826,\ PE^{1}$$

Treatment of Pupil-achievement Data

The remainder of this paper deals with some data[2] treated in the light of the Index of Integration.

Battery A, Form 1 of the Orlean's Achievement Test[3] was given, during the last two weeks of October, 1928, to 1239 school children including Grades III to IX. This test has five divisions including (*a*) reading, (*b*) arithmetic computation, (*c*) arithmetic reasoning, (*d*) language usage, and (*e*) spelling. For the present study the total individual-scores from all tests were used. It is seen that these are the tool subjects and the efforts made to partially integrate the pupils in these are to a degree justifiable. The primary data from this test are shown in Table I.

Table I.—The Primary Data from the Five Tests[1]

Grade	N	Range	Mean or average	SD ± PE
III	225	7–152	54.271	30.92 ± 0.98
IV	216	34–179	102.166	31.82 ± 1.03
V	208	60–261	147.586	32.18 ± 1.06
VI	204	85–292	185.534	30.56 ± 1.02
VII	208	110–308	222.110	34.39 ± 1.13
VIII	64	204–333	278.957	28.94 ± 1.73
IX	114	243–354	307.456	24.79 ± 1.10

[1] This table should be read: In Grade III there were 225 pupils who took the test making a range in scores from 7 to 152, with a mean score of 54.272 and a sigma of 30.92 ± 0.98.

[1] Kelley: "Statistical Method." P. 104.

[2] These data were secured through the courtesy of Superintendent J. N. Roeder, who gave half of the tests and supervised the giving of the remainder.

[3] Public School Achievement Test by Jacob S. Orleans, published (1928) by the Public School Publishing Company, Bloomington, Ill.

The population size in each grade gives the data some significance. The range of scores is more than one hundred in every grade, indicating that there are individuals in every grade who vary to a considerable degree in the number of learned pattern-responses. The data also portray an increasing number of learned pattern-responses with the advancing grades, as indicated by the mean values. The sigmas in each case are reliable as shown by their probable errors.

Statistical treatment of data takes for granted that its distribution closely conforms to the normal probability distribution. In order that we should not be "fooled by figures," the data were examined in this respect, the results being shown in Table II.

TABLE II.—THE SKEWNESS OF THE DATA[1]

Grade	N	Mean	Median	Skewness
III	225	54.271	48.17	+0.58
IV	216	102.116	102.20	−0.003
V	208	147.586	144.90	+0.25
VI	204	185.534	183.90	+0.16
VII	208	222.110	222.00	+0.009
VIII	64	278.957	280.90	−0.108
IX	114	307.456	308.40	−0.114

[1] Explanation: In the third grade there were 225 pupils who took the test scoring a mean of 54.271 and a median of 48.17, giving a distribution with a skewness of +0.58.

In no case was the skewness equal to one (1), indicating a satisfactory distribution.[1] The frequency polygons also pointed to this fact. The Grade VIII distribution, having a small population, produced a small assymetry.

The integration indexes were calculated for each grade as shown in Table III.

These indexes are comparable, permitting direct comparisons between different grades. It is seen, for example, that Grade VI is integrated more than three times that of Grade III. The Grade IX index is more than twice that of Grade VI, and more than six times that of Grade III.

---

[1] See Kelley: "Statistical Method." P. 62.

TABLE III.—GRADE INDEXES OF INTEGRATION AND YEARLY GAINS[1]

Grade	N	K	Yearly gains
III	225	1.75	
IV	216	3.209	1.46
V	208	4.58	1.37
VI	204	6.07	1.49
VII	208	6.46	0.39
VIII	64	9.63	3.17
IX	114	12.40	2.77
Mean yearly gain	...	.....	1.74

[1] Explanation: In Grade III, with a population of 225, the integration index was 1.75 while that of the Grade IV was 3.209 or a yearly increase of 1.46.

The yearly gains of one grade over another in their integration is rather uniform until the Grade VII increase is reached. Since

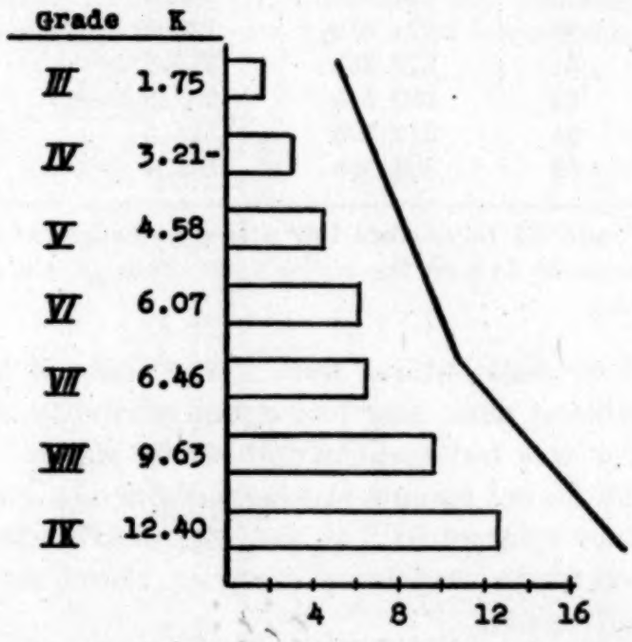


Fig.1. The Integration Indexes for Grades III to IX inclusive, determined from a total population of 1239 pupils. Note the accelerated integration of Grades VII, VIII, and IX, over Grades III to VI inclusive.

the test was given in October, this small increase may have largely occurred in Grade VI. The reason can only be guessed at. Data elsewhere scored on the Otis Classification Test reveal somewhat

similar results.[1] Grades VIII and IX seem to make up for lost time. The school system is doing its integration job to a partial extent at least.

The boys and girls were next compared as to their integration and mean yearly gains. The primary data are given in Table IV.

TABLE IV.—PRIMARY DATA FROM THE FIVE TESTS FOR BOYS AND GIRLS SEPARATELY[1]

Grade	Sex	N	Mean	Median	Skewness
III	G	103	60.252	54.8	+0.49
	B	122	49.221	42.65	+0.68
IV	G	102	104.176	105.8	−0.06
	B	114	100.368	96.4	+0.81
V	G	84	153.583	151.1	+0.24
	B	124	143.524	141.0	+0.23
VI	G	89	185.521	182.2	+0.32
	B	115	185.550	185.6	−0.004
VII	G	120	222.433	220.6	+0.18
	B	88	221.670	223.8	−0.13
VIII	G	31	277.289	275.8	+0.16
	B	33	280.515	287.8	−0.64
IX	G	51	312.705	314.5	−0.23
	B	63	303.206	302.4	+0.09

[1] Explanation: In Grade III there were 103 girls who made a mean score of 60.252 and a median score of 54.8 on the entire test. This gave a distribution with a skewness of +0.49.

The sigmas of these distributions were of the order of magnitude of those of the combined data, and had equal reliability since their largest probable error was but one-eleventh of its sigma. The distributions in Table IV do not reach a skewness of one (1), but by-and-large they are not as symmetrical as are the distributions of the combined data. Too much confidence dare not therefore be placed upon their subsequent results.

The following Table V shows the integration indexes for the girls and boys taken separately.

[1] These data were secured from a Vermont school and reveal the following K's for the similar grades:

III–1.5
IV–2.8
V–2.8
VI–4.8
VII–4.5
VIII–5.5
IX–7.9

Otis, A. S.: "Statistical Method in Educational Measurement." P. 276.

TABLE V.—INTEGRATION INDEXES FOR GIRLS AND BOYS SEPARATELY[1]

Grade	Sex	N	K	Kg-Kb
III	G	103	1.82	0.10
	B	122	1.72	
IV	G	102	3.58	0.53
	B	114	3.05	
V	G	84	5.11	0.75
	B	124	4.36	
VI	G	89	6.03	−0.07
	B	115	6.11	
VII	G	120	7.43	1.85
	B	88	5.58	
VIII	G	31	9.83	0.26
	B	33	9.57	
IX	G	51	13.33	1.21
	B	63	12.12	
Mean difference in favor of girls				0.66

[1] Explanation: In Grade III, with 103 girls, the integration index is 1.82, while that for the boys for the same grade is 1.72 or a decrease of 0.10.

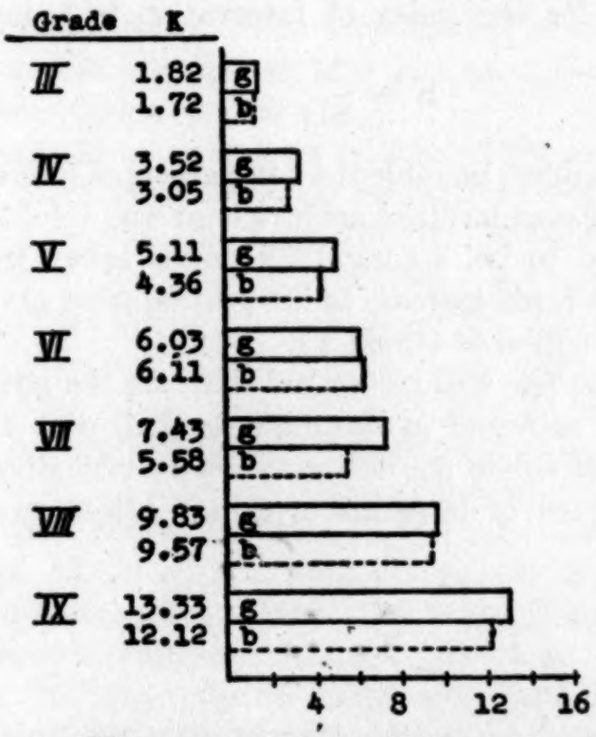


Fig. 2.—The Integration Indexes for boys and girls taken separately, as determined from a total population of 1239 pupils. Note that, with but one exception, the girls are integrated more than the boys.

Except for Grade VII, the girls are more highly integrated than are the boys, having a mean grade-difference of 0.66. The girls have a mean yearly gain of 1.92 as compared with 1.56 for that of the boys. An interesting fact is revealed about the Grade VII boys. They show a loss of 1.53 units from Grade VI to VII—about a year's loss. As was previously stated this may be a Grade VI loss.

Another interesting feature is found with the Grade VIII boys. Their yearly gain in integration reaches its climax in this grade—a figure of 3.99. Again the cause can only be guessed at. The yearly gains for the girls show a rather steady incline resulting in a greater ninth-grade integration than the boys possess. Apparently girls learn their pattern-responses more evenly, grade-for-grade, and cluster more about their mean value, than do the boys.

Conclusions

1. Pupil integration is being promoted by our schools.
2. The integration concept is placed on a quantitative, psychological basis which lends itself to statistical treatment.
3. The formula for the index of integration is found to be

$$K = \frac{M}{SD}$$

4. The formula renders possible the comparison of indexes revealed by all tests including standardized achievement and intelligence tests.
5. There is found to be, among 1239 pupils taken from Grades III to IX inclusive, a rapid increase in integration from grade to grade with the possible exception of Grade VI.
6. Boys tend to be less well integrated than are the girls in Grades III to IX inclusive, as found in the data studied, with the possible exception of Grade VI where the difference seems negligible.
7. The measurement of integration, calling attention to it, may add to its control.

SOME CONCLUSIONS FROM A COMPARISON OF SCHOOL ACHIEVEMENT OF CERTAIN RACIAL GROUPS

HELEN G. PRATT

Assistant Director of Research, Honolulu

Among the racial groups represented in Hawaii, the most interesting from the point of view of numbers involved and racial differences which it is assumed exist, are the Japanese, the Chinese, the Hawaiians and the part-Hawaiians. In a recent book Porteus and Babcock, of the University of Hawaii, have compared these groups in intelligence, in certain temperamental and social traits, and have quoted Stevenson Smith and Hoag on their school achievement. It is desirable that we have many more studies of school achievement which make use of more perfected tests as they are available. The writer has made a study of the achievement of representatives of the four social groups mentioned, and has used the Stanford Achievement Advanced Examination.

This examination was given in April 1924 and 1925 in Kaiulani School, and the Normal Training School. It was given in April 1924 in Kauluwela School, in April 1925 in Pohukaina School, and in January 1926 in Kamehameha Schools. All the schools except the last mentioned are public schools in Honolulu, and of the public schools, all except the Normal Training School are large schools, with enrolments of 1200 to 1700. Kamehameha Schools are private schools in Honolulu, and the enrolment in these schools is limited to those of Hawaiian blood. Ten of the 106 pure Hawaiian cases studied, and 68 of the 222 part-Hawaiian cases, came from Kamehameha Schools. All of the 12, 13, 14, and 15 year olds at the time of the tests were selected, and their scores tabulated by racial descent, according to the following groups—Japanese, Chinese, Hawaiian, and part-Hawaiian. There were 702 Japanese in these age groups, 351 Chinese, 106 Hawaiians, and 222 part-Hawaiians. A part-Hawaiian is any child of Hawaiian blood and the racial combination most frequently represented is the Chinese-Hawaiian. The different ages were represented as follows:

	Per cent of 12 year olds	Per cent of 13 year olds	Per cent of 14 year olds	Per cent of 15 year olds
Japanese	32	28	25	15
Chinese	31	27	27	15
Hawaiian	24	24	35	17
Part-Hawaiian	34	28	21	17

It is clear that any study involving so small a number cannot lead to conclusions regarding racial differences. The writer feels, however, that so large a sampling of the school population should lead to some conclusions regarding the comparative achievement of these racial groups in the schools, particularly in the case of the 702 Japanese and 351 Chinese, represented by almost identical per cents of each age group.

Certain questions may be asked.

1. Is there a difference in achievement level among these racial groups?
2. Is the difference which may be found a real difference or is it due to chance?
3. Are the differences large or small?
4. Are there differences in variability?
5. Are there any facts common to all groups?
6. Are there any facts which justify curriculum changes?

The first question may be answered by a statement of the mean scores. The following table presents these facts:

TABLE I.—MEAN SCORES ON THE STANFORD ACHIEVEMENT ADVANCED EXAMINATION

Test	Japanese	Chinese	Hawaiian	Part-Hawaiian
1. Paragraph meaning	54.90	59.05	46.3	58.05
2. Sentence meaning	30.6	34.0	26.65	33.85
3. Word knowledge	40.0	44.05	35.9	42.5
Total reading	124.13	136.7	107.3	133.5
4. Arithmetic computation	111.0	113.6	90.4	102.3
5. Arithmetic reasoning	58.65	63.7	43.0	56.7
Total arithmetic	169.5	177.28	134.6	158.0
6. Nature study and science	34.5	37.8	26.3	38.85
7. History and literature	35.6	39.55	26.2	34.5
8. Language usage	22.1	23.8	16.85	23.3
9. Spelling	127.5	131.2	109.4	119.3
Composite	51.21	53.78	42.0	50.8

It will be noticed from Table I that the Chinese group has the high score in every test but test 6, where the part-Hawaiian group has the high score. The Hawaiian group has the lowest score in every test.

Second high place is divided between the Japanese and part-Hawaiians. The part-Hawaiians surpass the Japanese in all the reading tests and in test 8, while the Japanese surpass the part-Hawaiians in all the arithmetic tests, spelling, and composite score.

All these differences great enough to be called real differences, and not ascribable to chance? If the differences between the means was as great as three times the sigma of the difference, the difference was called real. When the difference between the means approached three times the sigma of the difference, the Kelley-Wood table was used to find out how many chances to 1 there were that the difference was a real one. When the difference between the means did not approach three times the sigma of the difference, the difference was considered due to chance. Tables II, III and IV show the comparisons.

TABLE II.—CHINESE MEANS COMPARED WITH THOSE OF OTHER RACIAL GROUPS

Test	With Japanese			With part-Hawaiian			With Hawaiian		
	Difference between means	Real difference		Difference between means	Real difference		Difference between means	Real difference	
		Yes	No		Yes	No		Yes	No
1	4.15	x	.	1.00	.	x	12.75	x	
2	3.4	x	.	.15	.	x	7.35	x	
3	4.05	x	.	1.55	.	x	8.15	x	
Total reading	12.57	x	.	3.2	.	x	29.4	x	
4	2.6	.	x	11.3	x	.	23.2	x	
5	5.05	x	.	7.0	x	.	20.7	x	
Total arithmetic	7.78	499 to 1		19.28	x	.	42.68	x	
6	3.3	x	.	−1.05	.	x	11.5	x	
7	3.95	x	.	5.05	x	.	13.35	x	
8	1.7	142 to 1		.5	.	x	6.95	x	
9	3.7	.	x	11.9	x	.	21.8	x	
Composite	2.57	249 to 1		2.98	.	x	11.78	x	

We see from Table II that in the case of the Chinese and Japanese, the higher scores of the Chinese represent a real difference or a dif-

ference so great that the chance factor is negligible, except in the case of arithmetic computation and spelling. The only real differences between the Chinese and part-Hawaiians occur in arithmetic and in tests 7 and 9. The other differences are too small to be called real. The Chinese means are really higher than the Hawaiian means in every case.

TABLE III.—JAPANESE MEANS COMPARED WITH THOSE OF OTHER RACIAL GROUPS

Test	With Chinese			With part-Hawaiian			With Hawaiian		
	Difference between means	Real difference		Difference between means	Real difference		Difference between means	Real difference	
		Yes	No		Yes	No		Yes	No
1	−4.15	x	.	−3.15	110 to 1		8.60	x	
2	−3.4	x	.	−3.25	x	.	3.95	x	
3	−4.05	x	.	−2.5	199 to 1		4.1	.	x
Total reading	−12.57	x	.	−9.37	x	.	16.8	x	
4	−2.6	.	x	8.7	x	.	20.6	x	
5	−5.05	x	.	1.95	.	x	15.65	x	
Total arithmetic	−7.78	499 to 1		11.5	x	.	34.9	x	
6	−3.3	x	.	−4.35	x	.	8.2	x	
7	−3.95	x	.	1.1	.	x	9.4	x	
8	−1.7	142 to 1		−1.2	.	x	5.26	x	
9	−3.7	.	x	8.2	.	x	18.1	x	
Composite	−2.57	249 to 1		.41	.	x	9.21	x	

The part-Hawaiians are really superior to the Japanese in all the reading tests and in test 6. The Japanese are really superior to the part-Hawaiians in arithmetic computation. All the other differences are slight. In general the two racial groups have about the same achievement level, though they differ in the types of work in which they excel. The Japanese are superior to the Hawaiians in every test except the test of word knowledge. Here the difference lacks significance.

The part-Hawaiian group is really superior to the Hawaiian in every test but spelling, where the difference is not significant.

A comparison of the achievement of the Hawaiian group and all other groups may be found from Tables II, III and IV. This group

falls below the other groups and the difference is a real difference in practically every case and every comparison. The Hawaiian group of one hundred six cases is so small that the test results lack significance, but they may be taken as indicative of what probably would be found to be true if more cases could have been studied.

TABLE IV.—PART-HAWAIIAN MEANS COMPARED WITH THOSE OF OTHER RACIAL GROUPS

Test	With Chinese			With Japanese			With Hawaiian		
	Difference between means	Real difference		Difference between means	Real difference		Difference between means	Real difference	
		Yes	No		Yes	No		Yes	No
1	−1.00	.	x	3.15	110 to 1		11.75	x	
2	−.15	.	x	3.25	x	.	7.2	x	
3	−1.55	.	x	2.5	199 to 1		6.6	x	
Total reading	−3.2	.	x	9.37	x	.	26.2	x	
4	−11.3	x	.	−8.7	x	.	11.9	x	
5	−7.0	x	.	−1.95	.	x	13.7	x	
Total arithmetic	−19.28	x	.	−11.5	x	.	23.4	x	
6	1.05	.	x	4.35	x	.	12.55	x	
7	−5.05	x	.	−1.1	.	x	8.3	x	
8	−.5	.	x	1.2	.	x	6.45	x	
9	−11.9	x	.	−8.2	.	x	9.9	.	x
Composite	−2.98	.	x	−.41	.	x	8.8	x	

It has been shown that there are real differences in achievement level among the four racial groups under discussion. Are these differences large or small? In order to answer this question, the mean score on each test has been converted into its equivalent grade or G score. Table V shows the figures.

The smallest differences are, as has been shown, between the Chinese and part-Hawaiians. Here the differences in G scores range from no difference at all to .8, and a mean difference of .3. In terms of school grading, the difference between these groups is not small. Between the Chinese and Japanese, the difference ranges from .2 to .6, and the mean is .35. The greatest differences are between the Chinese and Hawaiian groups. The range is from .8 to 1.4, and the mean difference is 1.1. In general, we may conclude

that the differences among the racial groups, in terms of school grading, are not small and are largest when the Hawaiian group is compared with another group.

TABLE V.—MEAN G SCORES ON THE STANFORD ACHIEVEMENT ADVANCED EXAMINATION

Test	Japanese	Chinese	Hawaiian	Part-Hawaiian
1	5.3	5.7	4.5	5.7
2	4.9	5.3	4.4	5.3
3	5.7	6.1	5.3	6.0
Total reading	5.3	5.7	4.7	5.7
4	6.4	6.7	5.4	5.9
5	5.8	6.2	4.8	5.7
Total arithmetic	6.2	6.4	5.2	5.8
6	5.7	5.9	5.0	6.0
7	7.0	7.4	6.0	6.8
8	6.0	6.4	5.1	6.3
9	7.1	7.2	6.3	6.7
Composite	6.1	6.4	5.3	6.1

Are there differences in variability? Table VI shows the figures.

TABLE VI.—STANDARD DEVIATIONS, STANFORD ACHIEVEMENT TEST
Racial Descent

Test	Japanese	Chinese	Hawaiian	Part-Hawaiian
1	17.5	17.7	17.75	17.1
2	12.8	13.7	13.2	14.1
3	12.2	13.7	11.0	12.7
Total reading	38.6	41.8	38.0	41.0
4	24.0	24.8	25.1	27.6
5	22.5	21.3	19.6	22.5
Total arithmetic	43.7	42.1	40.0	47.1
6	17.6	17.5	16.7	19.1
7	19.5	20.4	15.7	18.6
8	10.5	10.9	10.1	11.3
9	37.6	41.0	41.6	42.7
Composite	14.83	15.2	14.8	16.3

The variability of the part-Hawaiian group tends to be greater than the variability of other groups. The variability of the Hawaiian group tends to be less than the variability of other groups. Porteus

and Babcock note the same tendency in the traits they measured. They also conclude that the Chinese group is slightly superior in general intelligence to the Japanese group, that the school achievement of the Chinese is slightly superior to the school achievement of the Japanese group.

Are there any facts common to all groups? A study of Table V shows: (1) The low scores for all groups are in reading. (2) Of the reading tests, the sentence meaning test shows the lowest score for every group, and the paragraph meaning has the next to the lowest score for every group. (3) Highest scores for the Japanese and Hawaiians are in spelling, for the other two groups, in test 7. (4) Second highest score for the Japanese and Hawaiians are in test 7; for the two other groups, in spelling. (5) Third high score occurs in arithmetic computation in the case of every group except the part-Hawaiian.

We may conclude from the above that the pupils in the schools of Hawaii, regardless of their racial descent, or their native ability, are noticeably backward in reading achievement, when this is compared with their achievement in other school subjects. It is evident that less effective teaching is being done in reading, and the more effective teaching occurs in spelling and arithmetic computation, two subjects of the rote memory type. It may also be concluded that there is such a thing as a " language handicap," and that it operates to lower reading achievement below achievement in the other school subjects. It is plain that the school environment has not operated to eliminate this handicap in the case of children 12–15 years old. A low reading level is characteristic of pupils in Grades IV–VIII and will probably be characteristic of them if we find them in the high school or university.

Inability of the Hawaiian group to adjust itself to the conventional school learning situation is apparent, though, as has been said, the numbers tested are so small that only tentative conclusions may be drawn. The pure-Hawaiian pupils are older, and their achievement materially lower than is the case with the other racial groups. It is plain that these other racial groups fit into the traditional school situation more successfully than the pure-Hawaiian. Every one engaged in teaching in the Islands is acquainted with the typical Hawaiian "misfit." He is older than the other pupils in the grade, big, nice, pleasant and agreeable, indolent, inefficient in school, and drifts out of it sooner or later into the streets. It is common enough in the Islands to speak sentimentally of the Hawaiian—of his good nature,

friendliness, and social qualities—and just as common to try to force him into a mold by which he is, by native ability and by interests, completely unfitted. If injustice is being done to any particular racial group in the schools, it is to the Hawaiian. Larger per cents of the Hawaiians are probably retarded, and leave school after having experienced non-promotion and failure. This is, of course, true of all the pupils of low school achievement, and curriculum changes are needed to take care of the "misfit" of any racial descent, but particularly to enable the schools to deal more successfully with the Hawaiian. The Hawaiian, or the retarded pupil of any racial descent, is not the misfit. Buckingham says that there cannot be a misfit child but that there can be misfit teachers and misfit courses of study. Curriculum changes are called for in Hawaii, that the large per cents of retarded Hawaiians and smaller per cents of retarded pupils of other racial descent find themselves no longer "misfits."

Summary

1. There is a real and not a small superiority in school achievement of Chinese over other groups.

2. Japanese and part-Hawaiians reach about the same level in school achievement, though they excel along different lines.

3. The Hawaiian falls below the other groups. It is suggested that this is proof of the existence of an unsuitable school environment and the need for curriculum changes.

4. All groups show low reading achievement. Special attention and changed emphasis needs to be given to reading. It is probable that there is such a thing as a language handicap, and that it is not overcome in Grades IV–VIII.

5. The present curriculum and teaching obtains greatest success in spelling and arithmetic computation.

A NOTE ON THE CORRECTNESS OF CERTAIN ERROR FORMULAS

KARL J. HOLZINGER

University of Chicago

In an article by the same title in the September issue of this journal, Dr. Harl R. Douglass has investigated the relative accuracy of two error formulas, and has offered a new one of his own. The object of this note is to comment further on the correctness of all of the above formulas.

The expansion suggested by Dr. W. E. Milne in the article quoted above is better written in the form

$$\sigma^2_{r_{nn}} = n^2[1 + (n-1)r]^{-4}[\mu_2 - 2k\mu_3 + 3k^2\mu_4 \cdots] \qquad (1)$$

where $\mu_i = \Sigma\delta^i$ divided by the number of samples.

If the odd moments are neglected in this series and even moments beyond μ_4 are also dropped, the remaining two terms in the bracket on the right may be simplified by assuming that $\beta_2 = 3$. Thus

$$\sigma_{r_{nn}} = n\sigma_r\,[1 + (n-1)r]^{-2}[1 + 9k^2\sigma_r^2]^{\frac{1}{2}} \qquad (2)$$

If $9k^2\sigma_r^2$ is negligible, formula (1) reduces to Dr. Shen's formula.

$$\sigma_{r_{nn}} = n(1-r^2)\,[N^{\frac{1}{2}}(1 + \{n-1\}r)^2]^{-1} \qquad (3)$$

which is incorrectly printed in Dr. Douglass' article. There is also an error on p. 435, in writing $n<1$ instead of $n>1$.

In the last part of his paper, Dr. Douglass gives a new formula which he says "will furnish values which, in at least most situations to which the Spearman-Brown formula would be applied" gives a still closer approximation than (3), or a formula by the writer. These remarkable statements are not only entirely unsupported, but the method used in the proof is inapplicable, the application made has a serious algebraic error, and finally the result is said to approximate the worst formula of the lot.

Dr. Douglass begins his proof by taking "as an approximation to $\sigma_{r_{nn}}$ the error in r_{nn} resulting from an error in r equivalent to σ_r." This he proceeds to do and arrives at his formula (6) which even with the above assumption is wrong. The factor preceding $(n-1)^2$ should be $(1-r^2)^2N^{-\frac{1}{2}}$ instead of $(1-r^2)N^{\frac{1}{2}}$.

The fault in Dr. Douglass' method is that it rests upon taking an arbitrary multiple of the standard deviation. If such a method is valid, it should hold for $t\sigma_r$ and $t\sigma_{r_{nn}}$ as well as for σ_r and $\sigma_{r_{nn}}$. Using the former values, we may write

$$r_{nn} + t\sigma_{r_{nn}} = n(r + t\sigma_r)[1 + (n - 1)(r + t\sigma_r)]^{-1}$$

and

$$r_{nn} - t\sigma_{r_{nn}} = n(r - t\sigma_r)[1 + (n - 1)(r - t\sigma_r)]^{-1}$$

Combining and simplifying these last two equations gives

$$\sigma_{r_{nn}} = n(1 - r^2)[N^{\frac{1}{2}}\{1 + (n - 1)r\}^2 - t^2(n - 1)^2(1 - r^2)^2N^{-\frac{1}{2}}]^{-1} \quad (4)$$

For t = 1 formula (4) reduces to

$$\sigma_{r_{nn}} = n(1 - r^2)[N^{\frac{1}{2}}\{1 + (n - 1)r\}^2 - (n - 1)^2(1 - r^2)^2N^{-\frac{1}{2}}]^{-1} \quad (5)$$

which is the correct form for Dr. Douglass' formula (6).

As an illustrative example, we may take as an approximation to $PE_{r_{nn}}$ the error in r_{nn} resulting from an error in r equivalent to PE_r. This amounts to letting $t = .6745$, and the resulting value from (4) is $\sigma_{r_{nn}} = n(1 - r^2)[N^{\frac{1}{2}}\{1 + (n - 1)r\}^2 - (.6745)^2(n-1)^2(1 - r^2)^2N^{-\frac{1}{2}}]^{-1}$.

Dr. Douglass' corrected formula (5) is thus clearly incorrect, because it depends upon the unit in which the error functions are measured.

The writer hoped to obtain a better formula than Dr. Shen's (3) when he gave the result listed as formula (2) in Dr. Douglass' paper. The formula as printed in the 1925 article is obviously wrong, and the error was due to carelessness of some sort for which the writer assumes all blame.

Dr. Shen's formula (3), or the slightly longer new formula (2) here proposed by the writer appear to be the best yet available for the value of $\sigma_{r_{nn}}$.

INCIDENTAL LEARNING[1]

RAYMOND R. WILLOUGHBY

Clark University

The Group of Subjects Used

The subjects were the first three hundred individuals tested in the experimenter's larger study[2] on mental inheritance. Since the families first tested were selected from the entire experimental group, because of geographic accessibility and other considerations not relevant to the test itself it will be sufficient to indicate how the basic group was chosen. The basic group comprised all children in the Palo Alto and Mayfield public schools born within a given year, the limits of the year being so chosen that a child born in the middle of it would be just thirteen years old at the estimated middle of the testing period; *i.e.*, the children were 12-6 to 13-6 as nearly as could be secured. An effort was made to test all these (*basics*), all their fathers and mothers, all brothers and sisters (*sibs*), all other available relatives (*remotes*) and any other persons sufficiently interested in the experiment to be willing to act as subjects (*irregulars*). Deductions from the group resulted from removals to inaccessible areas, not probably a selective factor; refusals, possibly a selective factor balancing or tending to balance the superior selection in this region, since the impression remains that more difficulty was experienced, on the average, in securing the cooperation of university professors' families, etc., than with ranchers, laborers, merchants, etc.; and a few cases in which the supposed basic was found to be more than thirteen years old. Many of the irregulars, on the other hand, were university people, some selected particularly for their known high ability. In addition, individuals who did not attempt recall on the substitution test were not included, whatever their performance on the substitution. No selection is presumably to be encountered on the ground of illiteracy in English or of inferior education, since the test appears to be almost completely non-verbal.

The females outnumber the males about five to four (167 to 133). This reflects (1) the preponderance of girl students on whom the test

---

[1] Acknowledgment is made to Dr. W. R. Miles for the fundamental suggestion for collecting the recall data.

[2] Genetic Psychology Monographs, 1927, Vol. II, 235–277.

was tried out for time, (2) the greater willingness of mothers to be tested, and (3) the fact that a considerable number of fathers of the basics were found to have died. The age distribution is as follows:

AGE DISTRIBUTION

Age	6-7:11	8-9:11	10-11:11	12-13:11	14-17:11	18-27:11	30-35:11
Frequency	5	18	21	80	37	15	17
Age	36-39:11	40-43:11	44-49:11	50-51:11	52-59:11	60-67:11	68+
Frequency	22	22	29	12	11	5	1

The children are evidently in the majority but there are almost as many adults, more spread out as to age, of course. The amount of education is from four grades to graduate degrees for the adults, and from none to the university for the children. The same is true of occupational level—on one end of the scale are strawberry-pickers and pick-and-shovel laborers, on the other university professors, army officers, etc. The impression is, however, that the peak is probably toward the high end in both these latter respects. The nationality distribution shows a marked predominance of a few standard stocks—English, Scotch, Irish, German, American. Italians are represented by two families, negroes by two, French by one or two, Spanish by a few, Japanese by one basic, etc., but the races of northwest Europe, more or less mixed, form the bulk of the group. This is roughly as would be expected in California. In short, the subjects are fairly representative of the general population of this vicinity, somewhat selected for superiority if anything.

METHODS AND CONDITIONS OF THE STUDY

After their selection in accordance with the methods already described the families of the regular group were visited by a field worker chosen for her tact, intelligence and agreeable personality, and persuaded, if possible, to make an appointment for an evening shortly to follow. On the appointed evening, either the field worker or the experimenter appeared, and after a few words of friendly reassurance and explanation as to the ease and agreeableness of the test

and its purpose, proceeded to the testing. So far as the material environment is concerned, therefore, the conditions are highly variable; the same is partly true of the social environment, but so far as possible every endeavor was made to keep the conditions in this sphere orderly and uniform. It is inevitable, of course, even in individual testing in a laboratory, that some subjects should be unruly, noisy, garrulous, inconsiderate of the rights of others, inattentive, stupid, etc., and with testing in homes these factors are all present, together with many others caused by the presence of a stranger and the novel situation. Every effort was made, however, in particular to keep all subjects interested and happy, whatever else happened; and it is believed that the invalidating factors have been limited to two or three, *viz.*, the tendency to "beat the time" by starting early, stopping late, or both (chiefly manifested by fathers worried over their family supremacy), surreptitious copying by small children, and distractions in the way of still smaller children who occasionally insisted on getting in the way of older subjects. These were allowed for whenever they were observed and could not be stopped without creating an embarrassing situation.

Although the tests were usually given in order (this order having been determined on as the best all-round) whenever a language or a temperamental situation required special treatment the order was changed as thought best. In most of these changes, Test 8, the one under consideration, was given first, for the reason that every subject could make a score, could understand readily what was desired, and would be likely therefore to gain confidence and familiarity with the general procedure at the start; this factor is not believed to introduce any important element of incomparability.

The time for Test 8 was two minutes, 30 seconds, which is somewhat too long, as evidenced by the number of subjects finishing. When this was discovered, however, it was too late to change it. The directions were very simple, comprising merely a reading of the line: "Put under each number the sign you find under it in the key above" and a supplementary illustration (if necessary) of the form "You see, here's a 3; all you need to do is just to look up here at the top to find what little sign goes with the 3, and put it *there;* see?" At the expiration of the time, and after all the usual comparisons of progress as between the subjects had been indulged in, the subjects were asked to turn their booklets over and fold the page back, so that they could not see what they had just done; then they were told to

write the numbers in the margin anywhere, and under each to write the sign that went with it in the test they had just finished—"as many as you can remember without looking back; and you have all

TEST 8.

1	2	3	4	5	6	7	8	9
−	∩	⊐	L	∪	o	∧	X	=

KEY

Put under each number the sign you find under it in the key above.

1.	3	1	2	1	3	2	1	4	2	3	5	2	9	1	4
2.	6	3	1	5	4	2	7	6	3	8	7	2	9	5	4
3.	6	3	7	2	8	1	9	5	8	4	7	3	6	9	5
4.	1	9	2	8	3	7	4	6	5	9	4	8	5	7	6
5.	9	3	8	6	4	1	5	7	2	6	2	4	8	1	3
6.	4	9	5	1	7	5	2	6	9	3	7	8	4	1	8

the time you wish;" with a few words of encouragement if any subject seemed discouraged by the task. With small children, the experimenter customarily wrote the numbers himself, saying "Now, do you remember what went with the 1? All right, you put it down

here." When all had remembered as many as possible, they were permitted to look back if they wished, and then proceeded to the next test.

Analysis of the Test

The test used in this section of the main investigation is the Digit-Symbol test from Army Beta. It contains ninety elements, arranged in six rows of fifteen elements each. Though the natural assumption would be that each symbol would therefore be practised ten times, such is not the case. The first four symbols are practised eleven times each, the five ten times, and the remaining four nine times each; the reason for this is not apparent. The table of practice for subjects finishing various fractions of the test is as follows:

Subjects Finishing *X* Elements Have Practiced *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i* times on 1, 2, 3, 4, 5, 6, 7, 8, 9 Respectively:

X	*a*	*b*	*c*	*d*	*e*	*f*	*t*	*h*	*i*
5	2	1	2						
10	3	3	3	1					
15	4	4	3	2	1				
20	5	4	4	3	2	1			
25	5	5	5	3	2	2	1	1	1
30	5	6	5	4	3	2	2	1	2
35	5	7	6	4	3	3	3	2	2
40	6	7	6	5	4	3	3	3	3
45	6	7	7	5	5	4	4	3	4
50	7	8	8	5	5	4	4	4	5
55	7	8	8	6	6	5	5	4	6
60	7	8	8	7	7	6	6	5	6
65	7	8	9	8	7	7	6	6	7
70	8	9	9	8	8	8	7	6	7
75	9	10	10	9	8	8	7	7	7
80	10	10	10	10	9	8	8	7	8
85	10	11	11	10	10	9	8	7	9
90	11	11	11	11	10	9	9	9	9

In general, it may be seen from the above that the first numbers are practised more at the start, the last at the finish, so that persons making a low score in the substitution function might also be expected

to make a low score in recall by reason of the lessened opportunity to learn the later elements; and they might also be expected to show a higher percentage of recalls on the first numbers. It will be shown repeatedly that the first statement does not hold; the second is treated under the head "Relation of Practice to Recall." The distribution of the different digits in the test may also be seen by comparing the intervals between two successive occurrences of the same digit:

Digit	Successive intervals											Mean interval
1	2	2	3	7	4	18	10	10	8	5	10	7.2
2	3	3	3	3	9	6	7	14	21	2	11	7.5
3	1	4	5	7	7	8	10	8	12	13	10	7.7
4	8	7	5	10	10	12	4	11	7	4	12	8.2
5	11	8	10	9	7	9	4	9	11	3	..	8.1
6	16	7	8	12	8	9	4	6	13	..	..	9.2
7	22	4	7	8	10	8	9	12	6	..	..	9.7
8	25	10	5	9	8	6	10	14	3	..	..	10.0
9	13	15	9	7	3	8	6	16	7	..	..	9.3

Substitution and Recall as Related to Sex, Age and Nationality

Sex.—The evidence indicates that, so far as the substitution ability is concerned, there are no significant sex differences. The constants are as follows:

	Males	Females
Mean (score points)	60.43, PE 1.01	62.33, PE 0.98
Standard deviation	17.53, PE .72	19.0 , PE 0.7
Variability (CV)	28.8	30.5

Difference between the means, 1.90; probable error of this difference, 1.41.

The chances are therefore about one in three that the true difference lies between the obtained difference and zero, and about one in five that the reverse is the true situation—*viz.*, that the males are higher

on the average than the females. The obtained difference in the measures of variability also lose their significance when it is learned that there were in the subject group fewer very young boys than girls; and at the other end of the scale, that a preponderance of the irregulars tested were girl students, selected for their known brilliance in order to try out the times and upper limits of the battery.

A similar condition obtains with reference to the recall ability. Although these data, on account of the truncation of the distributions and the small number of categories in which they are recorded, do not lend themselves readily to helpful interpretation through the calculation of constants, a glance at the accompanying histogram shows that the differences are entirely explicable by chance factors. The categories are numbers of symbols correctly recalled; the length of the bars are (males) numbers recalling each category and (females) this number multiplied by 0.8 to equalize the inequality in numbers (males 133, females 167).

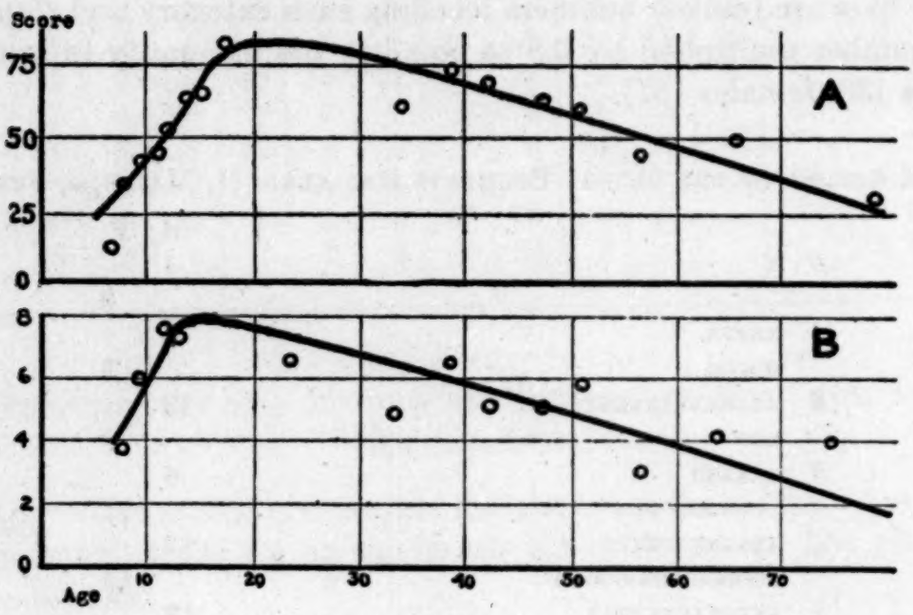
RECALL ABILITY OF THE SEXES—ELEMENTS RECALLED; (x, MALES; o, FEMALES)

		M	F
0	x	1	
			0
1	xxxxx	5	
	ooooo		5
2	xxxxxxxxxxxx	13	
	ooo		3
3	xxxxxx	6	
	oooooooooo		11
4	xxxxxxxxxxx	12	
	ooooooooooooooo		15
5	xxxxxxxxxxxx	13	
	oooooooooooooo		14
6	xxxxxxxxxxxxxxxxx	18	
	oooooooooooooo		14
7	xxxxxxxxxxxxxxxxxx	19	
	oooooooooooooooooooo		21
8	xxxxxxxxxxxxxxxxxx	19	
	ooooooooooooooooo		18
9	xxxxxxxxxxxxxxxxxxxxxxxxxxx	27	
	oooooooooooooooooooooooooooooo		31

Age.—With regard to the ability to substitute, the age factor seems to operate quite significantly. Reference to the graph on this point (Figure 1*A*) indicates the facts, which may be summed up

as follows: There is a strong tendency for the score means of the age arrays to lie on a Pearson Type III-i curve, having a mode in the neighborhood of twenty-two years and dropping away sharply toward childhood and gradually toward old age. The only considerable deviation from this occurs in the vicinity of thirty-two years, where there is a marked dropping off; it may be recalled, however, that the only subjects occurring in this region were the youngest parents of thirteen-year-olds, who were therefore married at twenty or before; and this group, as may be observed from the group in question as

Fig. 1. Trend of the means with age:
(A) Digit-symbol substitution ability;
(B) Ability to recall digit-symbol associations formed incidentally.



well as inferred on general principles, is an inferior one in general mental ability. Attempts have been made to discover some law governing the relation of variability to age, but these have not been successful. It is worthy of mention, however, that in eight of the sixteen age-arrays established and examined, the standard deviation does not vary significantly from 0.29 of the mean. Four of these occur in the region of thirty-six to fifty years, the same in which the means decline most regularly, and in which presumably the largest group of "normal" parents lie; so that in this range, at least, there is some reason to believe that both average ability and variation in this function decrease rather uniformly with increasing age.

The recall data present the same phenomena (Figure 1*B*), though in somewhat rougher form. Two factors are chiefly responsible for this roughness: (1) The smaller number of the score categories, now only nine in number, and (2) the pronounced massing of the scores at the high end, due to the ease of the task for most of the subjects. The latter factor particularly renders accurate estimation difficult, as means must be calculated on the assumption that the high end of the curve is essentially like the low. Due to these complicating elements no attempt has been made to estimate variabilities. As to means, however, it is worth noting that the same features are present as in the substitution data—the sharp initial rise, the mode in the vicinity of twenty years, and the gradual decrease to old age. Whether these result directly from the substitution curve itself, rather than being a corroborative measure in another ability, will be considered under the head of Relation of Practice to Recall.

Nationality.—The data on race and nationality are highly unsatisfactory, due chiefly to the difficulty of separating any one factor from the general complex for study. An individual of thirteen born in America of parents both of whom were born in one of the European countries whose populations are stabilized, is no doubt entirely comparable (sex differences being probably negligible) with another individual of thirteen born in America of parents born in another of such countries; but such individuals are few, and in order to secure enough of any group to be worth investigation at all from the present data, it has been necessary to compare several more or less incomparable groups. The factor of differing amounts of racial blood in individuals giving the same qualitative composition, and the intricacy of the qualitative compositions possible, have been summarily eliminated so far as possible (the raw data on such a point are always wrong or inadequate when collected by the questionnaire method) by considering only individuals giving but one strain, and all of such individuals; this causes the comparison, however, of persons of all ages, so that the soundness of the method depends on the soundness of the assumption that all ages are equally represented in each group. With large groups this is probably approximately true, but with small ones its truth nearly vanishes. The calculation of average deviations might be of service in estimating the comparative homogeneity of the groups, but this has not been thought worth while on account of the originally faulty nature of the data; for example, it is morally certain (though proof would be impossible) that the term "American"

represents a state of mind in the subject reporting rather than an objective fact—if indeed there be any fact to correspond to it.

Keeping this essential unreliability of the material in mind, then, means for each nationality group as above defined in substitution score and number of symbols recalled correctly have been computed, and are accompanied by the number of cases on which based:

Group	Cases	Mean score	Mean recall	Remarks
English	35	61.6	5.2	Possibly significant
Americans	20	62	5.6	Definition?
Scotch	9	56	6	High AD age
Dutch	9	69	7.7	Fairly homogeneous
Negro	8	48.5	5.5	Consistently low
German	7	63	6	High AD
Italian	7	61.3	7.3	Fairly homogeneous
French	5	51	5.3	Fairly homogeneous
Armenian	3	41	7.7	1 family—AD low
Greek	3	55	6.3	1 family—AD low
Swedish	3	71	7	AD fairly low
Irish	2	53.5	6	Two mothers much alike
Spanish	2	21.5	3.5	Girls thirteen and nine
Welsh	1	33	4	Man, low-grade
Danish	1	52	3	Woman, low-grade
Hungarian	1	82	3	Girl, thirteen
Japanese	1	71	9	Boy, thirteen

About all that can be deduced with any probability whatever from this table, then, is that the English group and the "American" group are markedly alike in this ability, having a mean somewhat below that of the basic 13-year-olds in substitution ability and substantially below them in ability to recall; differences observed between other nationality groups are of unknown significance.

The Relation of Practice to Recall

It has been noted (under "Analysis of the Test") that the amount of practice on different elements is different at different points in the test. It would be interesting to know, therefore, to what extent this inequality is a complicating factor in the general question. The increased and earlier practice on the first numbers does not, evidently, affect ability to recall them; and the effects of order of presentation

(greater ease of remembering the first and last elements of a series) are marked:

NUMBER OF TIMES SYMBOL FOR *X* WAS RECALLED

X	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
(Correctly).....	283	186	203	140	167	221	169	202	258
(Incorrectly).....	6	75	31	30	29	7	27	16	5

In compiling the above, an interesting fact regarding effect of age on recall ability in this connection was noted, viz., that the central "dip" in recall efficiency by number is more pronounced in the case of the older individuals:

RECALL EFFICIENCY BY AGE GROUP AND NUMBER, AVERAGES

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*0*
Basics.....	29	18	24	17	21	25	21	26	27	
Sibs.....	28	22	23	19	18	23	17	22	25	
Parents.....	27	17	14	7	12	18	12	16	25	
Mixed.....	28	17	21	13	17	23	19	19	26	

As an experiment in method, biserial coefficients of correlation between number of times practised and recall or non-recall were computed for digits 5 and 6 for 25 basics. The figures were .197, PE .158, and .109, PE .141, respectively, showing the absence of relationship as measured in this manner.

Since the biserial method would be exceedingly laborious if applied to each of the nine digits in turn, it was decided to compare the relations between score and practice on the digits between which practice was most divergent, in order to obtain a measure of the reliability of using the score as a measure of practice. The relation between score and practice on 1 (practice table, "Analysis of Test") is .979; that on 9, the most divergent from 1 as to practice, is .985; the difference, .006, is entirely negligible, and both are virtually perfect. An ordinary product-moment correlation between substitution score and recalled elements will therefore give (subject to the distorting influence of the truncation of the latter and partial truncation of the former) a very good measure of the relation of practice to recall. This coefficient is .369, PE .033, and demonstrates that there is a small but genuine relationship between the functions; this is in accord with the cruder results obtained elsewhere.

Substitution and Recall on Different Levels of General Mental Ability

It was chosen to make the study of this subject refer to "brightness" rather than "intelligence" (in Otis' sense, where the former corresponds to IQ and the latter to MA); therefore the subjects chosen from the group were all basics so far included in the files, to the number of 64, and the age factor is therefore eliminated, or at least confined within a narrow range. The best measure, obviously, is the product-moment coefficient, and the series to be correlated are total substitution score with total weighted score on the entire battery, and the latter with total elements recalled. The former gives a very satisfactory scatter, evenly distributed and with slight truncation; it yields

$$r_{\text{total score-substitution score, basics}} = .664, \text{ PE } .047$$

Taking the reliability coefficients for the battery and the test as .90 and .75 respectively, and correcting for the fact that the obtained correlation between measures is limited by the reliability of both (attenuation) we obtain a more probable coefficient of resemblance between substitution ability and general mental ability, as measured by this battery, of .80.

In recall, we are met again by the usual difficulty with truncation. The scatter, however, is large and irregular, and the coefficient is totally insignificant:

$$r_{\text{total score-recall elements, basics}} = .133, \text{ PE } .126$$

We are forced to conclude, then, within the limits of the data, that ability to recall does not depend on brightness, at least as measured by this battery in this age range.

(*To Be Concluded in the January Issue*)

FACTORS INFLUENCING RETENTION OF HISTORY IN THE SIXTH, SEVENTH AND EIGHTH GRADES[1]

SARAH JANET BASSETT

Baltimore Public Schools

What factors are closely related to ability to retain history in Grades VI, VII and VIII? The present paper reports data from testing 1364 pupils—729 boys and 635 girls—taught by twelve teachers in five public schools of Baltimore. The pupils represented wide ranges of intelligence levels, grade progress, chronological ages, interest and effort. The grade progress of these pupils was typical of large city school conditions, since 397 made rapid progress, 457 normal progress, and 510 slow progress.

History tests, devised by the writer were given at four-month intervals; namely, January, May, and September, 1926 to 1928. Six sets of tests were constructed—one for each of the following grades —VIB, VIA, VIIB, VIIA, VIIIB, and VIIIA. Each test was based on the material outlined in the History Course of Study in the City of Baltimore and comprised forty true-false, ten completion, twenty multiple-choice, and a choice of three out of four modified-essay questions. They were revised in the light of criticisms by several history teachers, supervisors, and administrators and of a preliminary try-out with 187 children.

A few of the questions for these grades are:

SIXTH GRADE[2]

True-false.—Spain did everything which she could to improve the conditions in her colonies.

Completion.—The_____ _____was written to keep the European nations from colonizing in America.

Multiple-choice.—Alfred improved his kingdom by (*a*) gathering all the best laws into one book, (*b*) fighting the Huns, (*c*) going on a Crusade, (*d*) teaching the French language.

Essay.—Why did the German barbarians succeed in conquering the Romans?

---

[1] For complete account of this investigation see: Bassett, S. J.: "Retention of History in the Sixth, Seventh, and Eighth Grades with Special Reference to the Factors That Influence Retention." The Johns Hopkins University Studies in Education, No. 12, The Johns Hopkins University Press.

[2] The first half of the questions given under each grade are for the lower section of the grade, that is, B section; while the multiple-choice and essay questions are for the higher or A section of a grade.

SEVENTH GRADE

True-false.—The Alien and Sedition Acts hastened the downfall of the Federalist party.

Completion.—The first Secretary of the Treasurer was_____.

Multiple-choice.—The last frontier was (*a*) along the Pacific coast, (*b*) along the Gulf of Mexico, (*c*) on the banks of the Mississippi River, (*d*) at the eastern base of the Rockies.

Essay.—How has the Industrial Revolution improved American life?

EIGHTH GRADE

True-false.—Westward expansion was one of the remote causes of the Civil War.

Completion.—The Compromise of_____proposed that California be admitted to the Union as a _____state.

Multiple-choice.—The Fordney-McCumber Tariff was a measure of (*a*) Democrats, (*b*) Roosevelt, (*c*) Republicans, (*d*) Socialists.

Essay.—How does the League of Nations seek to prevent war?

Not only were the children given tests to secure retention of history in real classroom situations, but wherever possible, the same children were given tests to supply the data for the influence of certain factors on retentiveness. Nine factors were considered; namely, (1) mental age, (2) chronological age, (3) subject preference, (4) interest and effort, (5) reading comprehension, (6) ability to answer questions on a standard history test, (7) weight, (8) height, and (9) sex.

In some cases mental age was secured by giving the McCall Multi-mental Scale, Form I, while in others, the Illinois General Intelligence Examination II, Form II, was given. The computations were treated on a basis of mental and chronological age, rather than as IQ's.

A rather simple device was employed to get the children's subject preference, for at this point in the experiment, the teachers of the various school studies came in to assist the securing of returns. The children listed on slips of paper the following: reading and literature, arithmetic, grammar, manual training (domestic science in the case of the girls), history, hygiene, geography, music, drawing, and physical training. Beside each study, the pupils wrote the numeral which represented how well they liked that particular subject. The one they liked the best of all received a score of ten, while the least preferred received only one.

The history teachers of the classes supplied the records for the children's interest and effort. The marks are quite significant, for they are composite scores based on how well the children paid atten-

tion, contributed to the history lesson, and participated in activities connected with history—in short, how much interested they were and how much effort had they put forth in the history classes. These interest and effort marks are closely allied to the term marks in history.

The reading comprehension scores were gotten by the use of the Monroe Standardized Reading Test Revised, Test II, or by the Chapman Unspeeded Reading Comprehension Tests (Grades V to XII). As a means of securing the ability to pass a standard history test, the children were given "Understanding of American History" by Pressey and Richards.

The weight and height scores were secured by the school nurses. Weight calculations were based on the actual weight divided by the normal weight as determined by the "Weight-height Tables for Boys and Girls," prepared by Bird T. Baldwin and Thomas D. Wood.

Three equated groups, totaling one hundred ninety children, supplied the data for the influence of sex on history retention. These were chosen from VIIA, VIIB, and VIA.

An exact method of ascertaining the relationship between retention of history and factors that may influence retention of history and factors that may influence retention was to find the correlations between the retention tests after different time intervals—on the initial test (T_0) after four (T_4), eight (T_8), twelve (T_{12}) and sixteen months (T_{16})—and the factors. The same data were used to calculate the intercorrelations between the factors themselves. These correlations were computed according to the Pearson Product Moment Coefficient of Correlation.

According to Table I, the mean of the correlations between retention and mental age is .366, with a range from .167 to .707. Throughout this study, consequently, a slight relationship exists between retention and mental age; that is, the children with the higher mental ages tend to remember better than those with lower mental ages.

The mean of the correlations between retention and chronological age is $-.303$, while the range for these correlations extends from $-.010$ to $-.569$. Hence, there is a low negative relationship between retentiveness and chronological age which signifies that, in general, the older children in our classrooms are doing less retaining of their history knowledge.

The Pearson r's between retentiveness and subject preference have a mean of .359 with a range of .001 to .611. From these data, it is safe to conclude that there are low, yet in some instances signifi-

cant, correlations between retention and preference for history; that is, the children who like history tend to retain it.

TABLE I.—MEANS AND RANGES OF THE CORRELATIONS BETWEEN HISTORY RETENTION AND THE FACTORS

	Means	Ranges
History retention and mental age	.366	−.617 to .707
History retention and chronological age	−.303	−.010 to −.569
History retention and subject preference	.359	.001 to .611
History retention and interest and effort	.664	.490 to .805
History retention and reading comprehension	.483	.245 to .677
History retention and Pressey-Richards history	.526	.238 to .719
History retention and height	−.098	−.395 to .319
History retention and weight	−.023	−.376 to .301

Table I shows that the highest mean correlation exists between retention and the factor, interest and effort, since it is .664 with a range from .490 to .805. Such measures of relationship make it possible to conclude definitely that a positive relation exist between ability to retain and the interest and effort which children put forth to master history. E. Meumann's expression bears out this conclusion, for he states, "Memory fails to retain many impressions that come to us countless times during our lives. We remember only what we have apprehended attentively and with the intention of remembering it."[1]

Ability to read tends to be an asset to remembering for a mean correlation of .483 with a range from .245 to .677 is found between retention and reading comprehension.

The second highest mean of the Pearson *r*'s exists between retention and the achievement scores on the Pressey-Richards History Test, for the mean is .526 with a range from .238 to .719. Probably the explanation of this positive relationship is found in the fact that the standard history test represents a fund of familiar information—the apperception mass. The greater the amount of this "old" knowledge, the better are the chances to associate the "new" history material with it. Such a relation is explained by Thorndike's sentence, "The

[1] Meumann, E.: "The Psychology of Learning: An Experimental Investigation of the Economy and Technique of Memory." Translated by John Wallace Baird, D. Appleton and Co., 1913.

progress of knowledge is far less a matter of acquaintance with more and more gross situations in the world than it is a matter of insight into the constitution and relations of long familiar ones."[1]

Little relationship exists between the remembrance of history and a physical measure, the height of children, for there is a slight negative mean correlation −.098, with a range from −.395 to .319. Practically the same condition is found between retentiveness and weight, since there is a mean correlation of −.023, with a range from −.376 to .301.

What is the influence of sex, the ninth factor, on the retention of history? The records for the one hundred ninety pupils, ninety-five boys and ninety-five girls from VIIA, VIIB, and VIA were equated on the bases of (1) initial retention history test (T_0), (2) interest and effort (IE), and (3) ability to pass the Pressey-Richards History Test (His). They are given in Table II. The factors, interest and effort and ability to pass the standard history test, were chosen as the bases for the equating because, in the light of the present investigation, these give the highest prediction for history retention. Table II shows the means and Q's for the three equivalent groups of boys and girls.

TABLE II.—GROUPS OF BOYS AND GIRLS EQUATED ON THE BASES OF THE INITIAL HISTORY TEST (T_0), INTEREST AND EFFORT (IE), AND ABILITY TO PASS THE PRESSEY-RICHARDS HISTORY TEST, "THE UNDERSTANDING OF AMERICAN HISTORY" (HIS)

	T_0	IE	His		T_0	IE	His
44 VIIA boys				44 VIIA girls			
M[1]	60.93	74.70	24.36	M[1]	60.45	75.27	23.63
Q	8.08	10.17	5.50	Q	8.00	9.03	4.59
23 VIIB boys				23 VIIB girls			
M	56.39	69.96	26.57	M	56.48	70.35	24.63
Q	12.25	14.75	5.80	Q	11.94	14.90	5.05
28 VIA boys				28 VIA boys			
M	53.61	74.54	26.75	M	53.55	75.04	26.18
Q	5.88	7.50	4.38	Q	5.67	7.62	4.00

[1] M = Mean.

The extent to which the equated groups of boys are superior to the groups of girls is given in Table III.

[1] Thorndike, E. L.: "Educational Psychology; Briefer Course." Teachers College, Columbia University, 1923, p. 148.

TABLE III.—INFLUENCE OF SEX RETENTION OF HISTORY SHOWN BY MEANS AND Q DIFFERENCES FOR EQUATED GROUPS OF BOYS AND GIRLS

Retention tests	T_0[1]	T_4	T_8	T_{12}	T_{16}
44 VIIA boys: mean	60.93	54.95	50.98	48.68	
44 VIIA girls: mean	60.45	53.34	48.68	47.36	
Difference in means	−0.48	−1.61	−2.30	−1.32	
Q difference	.....	.....	.....	1.392	
23 VIIB boys: mean	56.39	46.91	44.70	39.44	37.22
23 VIIB girls: mean	56.48	45.61	42.04	38.26	36.17
Difference in means	+0.09	−1.30	−2.66	−1.18	−1.05
Q difference	.....	.....	.....	2.386	
28 VIA boys: mean	53.61	46.21	43.68	42.18	
28 VIA girls: mean	53.55	44.93	42.39	39.07	
Difference in means	−0.06	−1.28	−1.29	−3.11	
Q difference	.....	.....	.....	1.746	

[1] T_0 = Initial Retention History Test.
T_4 = Test after interval of four months.
T_8 = Test after interval of eight months.
T_{12} = Test after interval of twelve months.
T_{16} = Test after interval of sixteen months.

On the four (T_4), eight (T_8), and twelve (T_{12}) month intervals the forty-four VIIA boys excel the mean history scores of the forty-four VIIA girls by 1.61, 2.30, and 1.32 points,[1] respectively. For the same time intervals and for the sixteen month lapse (T_{16}), twenty-three VIIB boys surpass the twenty-three VIIB girls by 1.30, 2.66, 1.18, and 1.05 points. The group of twenty-eight VIA boys, likewise, slightly excel the twenty-eight VIA girls by 1.28, 1.29, and 3.11 points for the four, eight, and twelve-month intervals, respectively. All the data show that the boys are superior to the girls in the ability to retain historical knowledge. Consequently, sex is a factor that tends to influence history retention.

The relation of two or more factors to the retention of history in Grades VI, VII and VIII was secured by the computation of coefficients of multiple correlations between the history tests and the different factors. The means and ranges of the multiple correlations are tabulated in Table IV.

A combination of interest and effort and ability to pass the Pressey-Richards Test with the criterion, history retention, gives a mean multiple correlation of .699 with a range from .520 to .822. Hence,

[1] Points = raw scores.

combining these two factors gives ability to retain. When subject preference is added to these two factors, the mean multiple correlations increases to .722 with a range of .581 to .834.

TABLE IV.—MEANS AND RANGES FOR MULTIPLE CORRELATIONS BETWEEN RETENTION AND THE FACTORS

	Means	Ranges
$R_{1.24}$[1]	.699	.520 to .822
$R_{1.245}$	.722	.581 to .834
$R_{1.243}$	.723	.522 to .853
$R_{1.2456}$	.737	.600 to .840
$R_{1.2453}$	.738	.598 to .860

"$R_{1.24}$ = mean of .699" signifies that the multiple correlation between retention tests and a combination of interest and effort and ability on the Pressey-Richards History Test averages .699, etc.

[1] 1 = History Retention Test.
2 = Interest and effort.
3 = Mental age.
4 = Pressey-Richards History Test.
5 = Subject preference.
6 = Reading comprehension.

The mean multiple correlation of interest and effort, ability on the Pressey-Richards History Test, and mental age is .723 with a range from .522 to .853. By combining interest and effort, ability on the Pressey-Richards History Test, subject preference, and reading comprehension, a mean correlation of .737 with a range from .600 to .840 is obtained. Such a combination of factors is predictive of successful history retention.

Of all the combinations of factors with the history retention the highest mean multiple correlation, .738, with a range from .598 to .860 is secured by combining interest and effort, ability on the Pressey-Richards History Test, subject preference and mental age.

CONCLUSIONS

1. The data of the present investigation show positive, though low, relationships between retention and mental age, subject preference, and reading comprehension.

2. Higher positive relationships exist between history retention and interest and effort and ability on the Pressey-Richards History Test, "The Understanding of American History."

3. The relationship between retention and chronological age is consistently negative.

4. For height and weight, the relationships are too low and variable to have predictive value.

5. Boys are slightly superior to girls in the ability to retain history knowledge; hence, sex is a factor that influences history retention.

6. A combination of interest and effort and ability to pass the Pressey-Richards History Test give multiple correlations ranging from .520 to .822, averaging .699.

7. When subject preference is added to this combination, the coefficients average .722 and range from .581 to .834.

8. If instead of subject preference, we add mental age the coefficients average, .723 and range from .522 to .853.

9. By combining interest and effort, ability on the Pressey-Richards History Test, subject preference, and reading comprehension, we get multiple correlations that average .737 and range from .600 to .840.

10. If we substitute mental age for reading comprehension, the coefficients average .738 and range from .598 to .860. Therefore, the combination of factors given in conclusions 9 and 10 give the highest predictive value for the retention of history in Grades VI, VII and VIII.

NOTE ON RELIABILITY PREDICTIONS FOR THE SEASHORE MUSIC TESTS

LYLE H. LANIER

Vanderbilt University

Recently I read the brief article by Farnsworth (this Journal, Vol. XIX, pp. 586–588) on the use of the Spearman-Brown prophecy formula for predicting the reliability of the Seashore music tests. Farnsworth purports to find discrepancies between his results and those earlier reported by me (*Journal Experimental Psychology*, Vol. X, pp. 69–113). As a matter of fact, no discrepancies exist, inasmuch as such reliability predictions in my paper as are comparable with his agree quite well with the actual reliabilities.

Farnsworth must have overlooked the parts of my article which reported comparisons of predicted with actual reliabilities secured by sub-dividing a single test (pp. 86–89). The principal conclusion from my studies was that fairly good agreement existed between actual and predicted reliabilities when one test only (*i.e.*, a single sitting) was used, whereas there was over-prediction, generally, when two separate testings were involved.

The following table will summarize the data reported by Farnsworth and by me for the tests in question:

	Pitch		Intensity		Time	
	Actual	Predicted	Actual	Predicted	Actual	Predicted
Farnsworth, r halves of one test..	.724	.729	.508	.561	.460	.537
Lanier, r halves of one test.....	.839	.834	.610	.613	.583	.570
Lanier, r separate tests.....	.680	.897	.597	.799	.495	.703

Three facts are apparent: (1) in every case but one (and that involves repeated tests) the actual reliabilities secured by Farnsworth are lower than mine; (2) my predicted reliabilities are almost identical with the empirical coefficients for predictions *within a single test*, the agreement being closer than Farnsworth obtains; (3) my actual reliabilities, secured by correlating repetitions of tests, are almost as high for pitch, and are higher for intensity and time than those

secured by Farnsworth with the split-half technique (which he tries to advocate as superior to the method of correlating repeated tests). It would seem then that his objection to my methods was pointless, since I used *both* means of getting actual reliabilities. Furthermore, even if I had restricted my work to the use of repeated tests, for checking the accuracy of the prediction formula, the reliabilities which I secured by this method are higher than the empirical coefficients Farnsworth obtained by correlating halves of single applications of the tests.

Not only is Farnsworth's proposed method of checking the Spearman-Brown formula lacking in originality, but also his discussion of possible causes of the unreliability of the Seashore tests, as determined by correlating repeated tests, will be found in slightly different form on pages 108 and 109 of my article.

REPLY TO DR. LANIER'S NOTE ON THE SEASHORE TESTS

PAUL R. FARNSWORTH

Stanford University

Dr. Lanier has taken me to task for my article in this Journal, Vol. XIX, pp. 586–588. As is usual in controversies of this sort, we are apparently not able to understand each other clearly. I did not attack Dr. Lanier's data. I merely gave added experimental material which led me to conclusions somewhat different from those of Dr. Lanier.

Let me quote from his note: "Note only is Farnsworth's proposed method of checking the Spearman-Brown formula lacking in originality but also his discussion of possible causes of the unreliability of the Seashore tests, as determined by correctly repeated tests, will be found, in slightly different form, on pages 108 and 109 of my article."

It seems to me that this statement needs slight overhauling. First let us consider the unoriginality of my check method. A careful perusal of the Lanier article clearly shows that Lanier himself did not use this method. And, since I know of no other such study, I would be inclined to call his statement an error. Lanier's predicted reliabilities as given in his note are derived from correlating rows 1 and 2, 2 and 3, and averaging the correlations. The Spearman-Brown formula is then invoked. My predicted reliabilities make use of a variety of split halves. Where possible, the odd-even items of the top five rows are employed. This method is checked by employing the bottom five rows as well. In the intensity test the arrangement makes it possible to split the record faces and to subdivide these. In the time record it was possible to create two tests forms by grouping certain columns.

As to my unoriginality in supplying reasons for the possible retest unreliabilities, I need only say that music testers have been long aware of the fatigue and ennui which frequently arise from retests of the Seashore records. At least this idea was in published form as early as Heinlein's 1925 article which preceded the work of Dr. Lanier. I would under no circumstances claim originality for such commonly held theories.

Notwithstanding these obvious difficulties with retests, Dr. Lanier bases his criterion of reliability on them. Note his statement as

listed under Predictions based on two applications of the Tests: "The results from the tests of pitch, consonance, and rhythm show that all three might be decreased in length at least one-half with no loss in reliability." This still appears to me as an unwise procedure.

I agree with Dr. Lanier's condemnation of the policy of substituting mathematical treatment of data for careful experimentation and control of conditions. However, from a consideration of the data appearing in the Lanier article and in my article, and from observations of more recent reliability work on numerous other groups, I believe that the Spearman-Brown formula functions as perfectly as can be hoped when reliabilities on groups agree no better than they do.

PHOTOGRAPHIC RECORDS OF EYE-MOVEMENTS IN STUDYING SPELLING[1]

ETHEL M. ABERNETHY

Queens College

The Problem

The main purpose of this study was to investigate the relation between the method of learning to spell, as indicated by photographic records of eye-movements, and efficiency in learning. Laboratory studies of the reading process have revealed the importance of eye-movements in the diagnosis of difficulties in reading. The experiment to be reported here was concerned with determining the applicability of this technique of the laboratory to the analysis of processes in spelling.

The attack upon this general problem was through a comparative study of good and poor spellers with regard to the following fundamental characteristics of eye-movements:

1. *The Temporal Factor.*—The time spent in study, the number of fixations and the average duration of the pauses in studying the word.

2. *The Unit of Study.*—Whether the study is principally by syllables or by the smaller unit of the letter.

3. *The Order of Fixations.*—Whether the eye-movements are predominantly progressive or regressive; whether the study is made primarily by surveys of the word as a whole or by analysis of one part at a time.

4. *Difficulty as a Factor in the Focalization of Attentions.*—The degree of concentration upon the difficult parts of the word.

Procedure

Apparatus and Technique Used in Securing Photographic Records.—A detailed account of the apparatus employed in this experiment has

---

[1] The experiment was conducted in the laboratories of the School of Education, the University of Chicago. The writer is indebted to Professors Freeman, Buswell and Breed, and to the officials of the University of Chicago Laboratory Schools, for their interest and cooperation in the investigation.

been given C. T. Gray.[1] Buswell,[2] in a later monograph, has described modifications in the equipment. The method of determining from the film record the location of the fixations on the printed line in reading has been fully described by Buswell.[3] The same method was used in locating fixations on the word in spelling.

The Subjects.—Twelve individuals served as subjects: Four good spellers and four poor spellers of Grade VI of the University of Chicago Elementary School, and four graduate students of the university. Selection of the Grade VI subjects was made on the basis of pupils' scores in this grade on the following tests: Seven class pre-tests given during the school year; lists M, O and Q of the Ayres Spelling Scale, and (for comparison with the adult subjects) the last five lists of the Buckingham Extension. After dividing the class into quintiles on the basis of these measures, four good spellers were chosen from the highest quintile and four poor spellers from the lowest. In selecting the subjects, all were eliminated who had any one of the following handicaps: Defective vision, defective hearing, IQ below 100, language difficulty due to foreign parentage. Moreover, none of the pupils who had received remedial treatment in spelling was employed. In IQ's the two groups were fairly comparable.

The adult subjects were members of classes in experimental education. Their average score upon the last five lists of Buckingham Extension was eighty per cent. For the good Grade VI spellers the average was 49.1 per cent. No word of this test was correctly spelled by any one of the poor spellers.

Materials and Method

1. *Sixth-grade Pupils.*—The list of ten words used for the Grade VI pupils was selected from words misspelled by all eight of the subjects on preliminary tests given within a few days of the experiment. In order that the material might be meaningful, commonly used words—

---

[1] Gray, C. T.: "Types of Reading Ability as Exhibited through Tests and Laboratory Experiments." Supplementary Educational Monographs, Vol. I, No. 5, pp. 83–90. Chicago: Department of Education, University of Chicago, 1917.

[2] Buswell, G. T.: "Fundamental Reading Habits: A Study of Their Development." Supplementary Educational Monographs, No. 21, 1922, pp. 12–13.

[3] Buswell, G. T.: "A Study of the Eye-voice Span in Reading." Supplementary Educational Monographs No. 17, 1920, pp. 4 and 5.

such as *accommodation, occasionally* and *appreciative*—were employed.[1] In the laboratory, the list of words was read and the subject was requested to state the meaning. The subject was then seated before the camera and the words were presented singly. The first two words were used as practice material while the subject was becoming accustomed to studying before the camera. The direction was to study each word until satisfied that it was mastered. To make the condition of study as normal as possible, the subject was told that a test would be given at the conclusion of the study period. As soon as the subject was removed from position before the camera, the words were dictated for writing.

2. *Adults.*—Ten technical words, selected from botany and chemistry, were used for the adult subjects. None of the subjects was familiar with any word in its complete form, but each recognized some part of nearly every word. The preliminary test was given immediately before the study period. The technique in the laboratory was the same for the adults as for the Grade VI subjects, except that with the adults it was necessary for the experimenter to state the meaning of the words.

Results

Time Spent in Study and Efficiency in Learning.—InTable I is presented a summary of group differences in learning time and in amount learned. The average study of the good Grade VI speller is seventy per cent of that of the poor; while the average time of adults, upon more

Table I.—Record of Learning

Subjects	Average number of fixations a word	Average learning time (units 1/25 second)	Average duration of fixations (units 1/25 second)	Average per cent of letters correct on pre-test	Average per cent of letters correct on final test
Adults	24.8	220.3	8.8	80.3	93.6
Good Grade VI spellers	28.1	385.8	13.7	85.5	95.6
Poor Grade VI spellers	44.8	550.5	12.3	76.1	77.6

[1] All the words presented in the experiment were printed in eighteen point type. Preliminary experimentation indicated that, within ordinary limits, the size of type does not influence the number of fixations.

difficult words, is only fifty-seven per cent of the learning time of the good spellers. When spelling efficiency is estimated in terms of the per cent of letters correctly given on each of the tests, the gain from study averages 13.3 per cent for adults, 10.1 per cent for the good spellers and only 1.5 per cent for the poor.

The average duration of the fixations is considerably less for the adults than for either group of the Grade VI subjects: in twenty-fifths of a second—adults, 8.8 ± .75; good spellers, 13.7 ± .7; poor spellers, 12.3 ± .5.

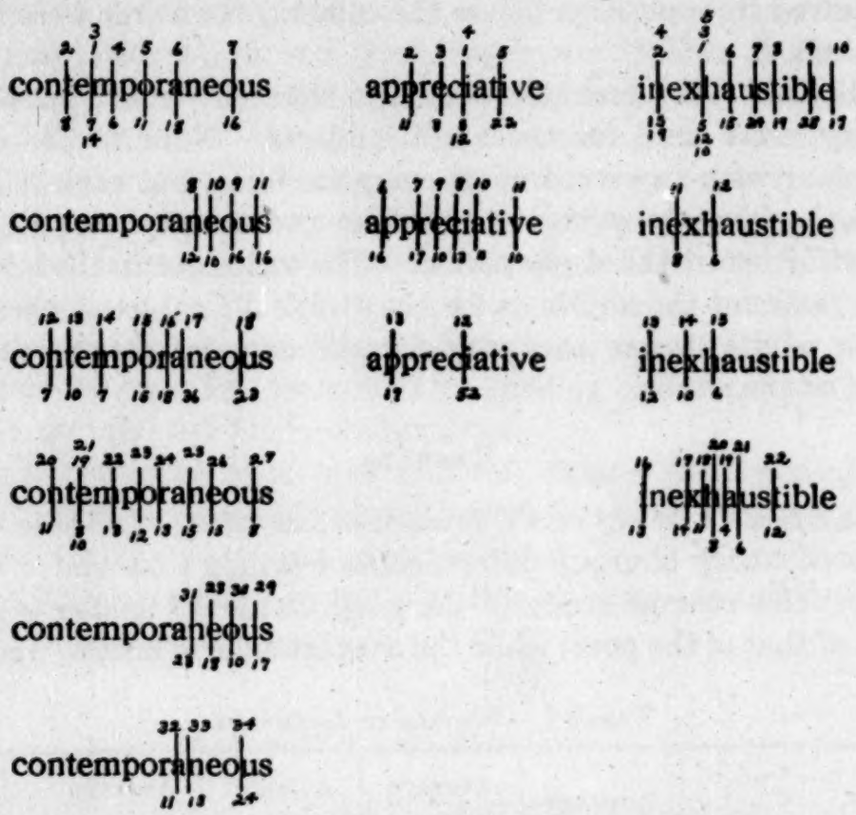


PLATE I.—From records of a good sixth-grade speller.

The Unit of Study.—Within each of the three groups, there are individual differences in the characteristic unit of study. It may be stated, however, that the adults employ, on the average, a broader recognition unit than either group of younger subjects. Of the four adult subjects, three study principally by syllables, one by letters. A comparison of the two Grade VI groups shows no relation between the breadth of the unit in study and efficiency in learning. In each group are subjects who study mainly by syllables and others who study letter by letter.

The Order of Fixations.—The adult subjects make comparatively few regressive eye-movements. The average per cent of backward movements is 11. ± 3.7 for the adults, 25.4 ± 1.1 for the good spellers, 23.5 ± 2.2 for the poor. (The return of the eye to the first of the word for the beginning of a new survey was not counted a regression.)

The adults study a word primarily by surveys of the word as a whole. Within each of Grade VI groups individuals differ widely in the characteristic order of fixations.

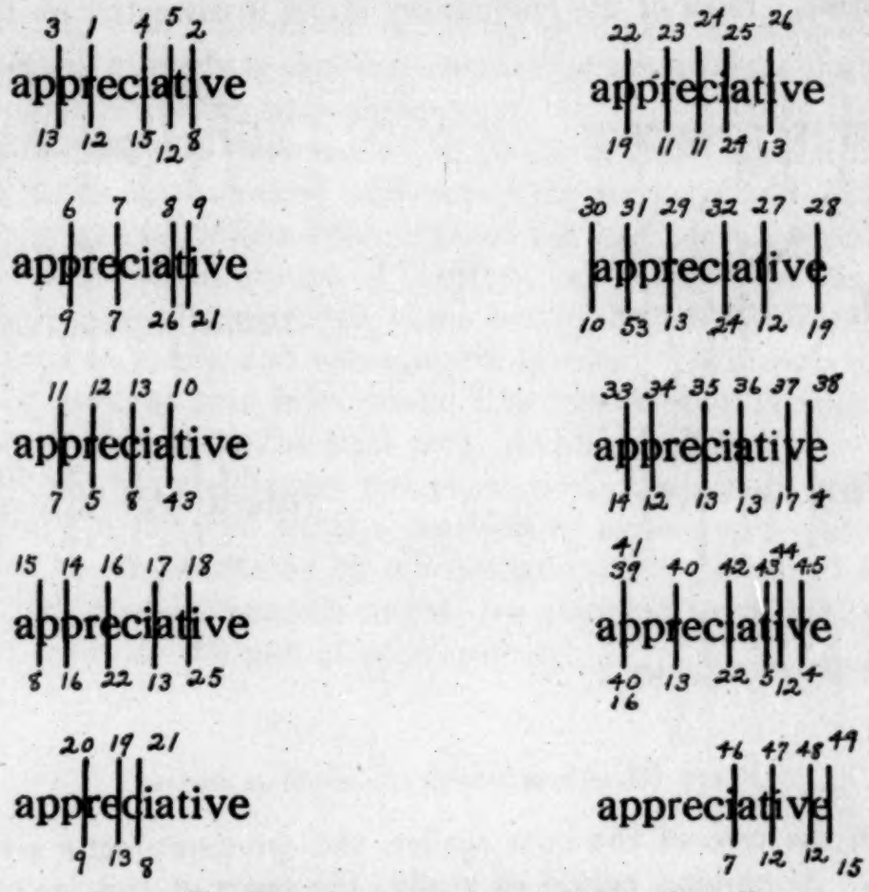


PLATE II.—From records of a poor sixth-grade speller.

Difficulties as a Factor in the Focalization of Attention.—The good and the poor spellers of Grade VI differ in method of study mainly in that the good spellers apparently recognize their errors and tend to make a systematic attack upon the difficult parts of the words, while the poor spellers seem to be slow in the recognition of difficulties. Plate I[1] shows typical records of the most efficient of the good spellers.

[1] The order of the fixations is indicated by the serial numbers above the vertical lines; at the lower end of each line the duration of the pause is given in twenty-fifths of a second.

Plate II is representative of the records of a poor speller. The good speller concentrates attention upon the difficult part of each word. For example, in *contemporaneous* (spelled *contemporanios* in the preliminary test) there is first a survey of the whole word, then a study of the difficult syllables, a further analysis of the entire word—two complete surveys—and a final detailed study of the troublesome parts. Records of the other words (first spelled *appreshiative* and *inexaustible*) also indicate an awareness of error and an orderly system of analysis. Each of the preliminary errors is corrected on the final

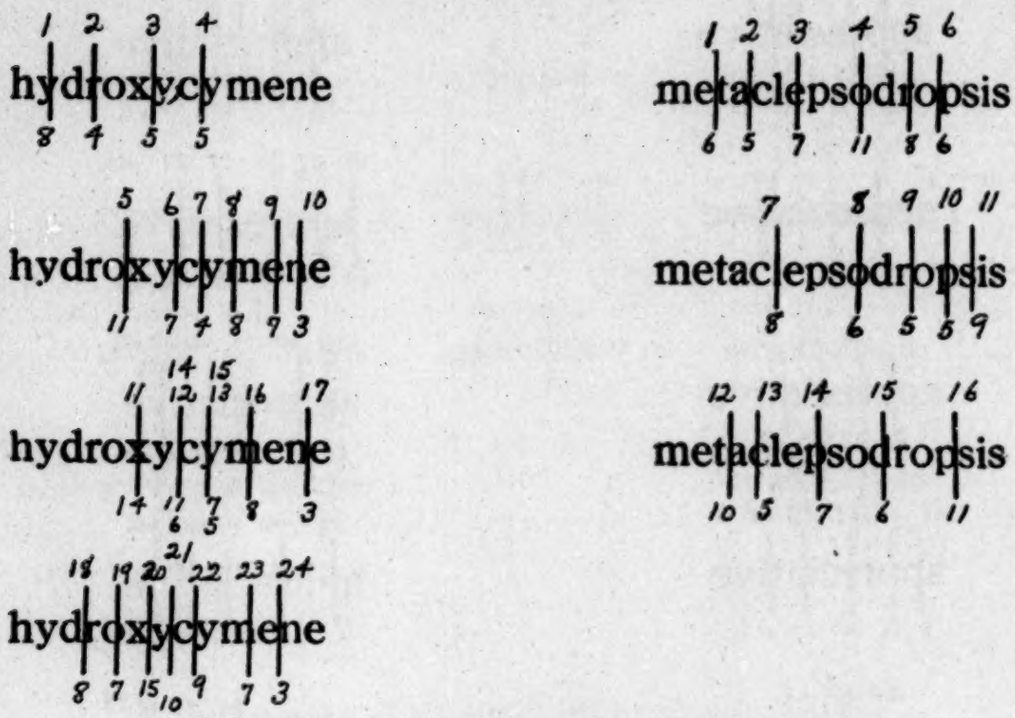



PLATE III.—From records of a graduate student.

test. In the case of the poor speller, the eye-movements are more random. After long period of study, the error of the preliminary test, the omission of the letter *c*, is uncorrected.

The records of Plate III are selected from those of the most efficient adult subject. This subject makes a detailed analysis only of difficult parts of a word. Only about two per cent of his eye-movements are regressive, and the regressions occur mainly where there is special difficulty. (Preliminary spelling, *hydroxapsychemene* and *metaklypsodropsis*.)

SUMMARY OF RESULTS

1. Comparison of four adult subjects with eight subjects of Grade VI, four good spellers and four poor spellers, in photographic records of

eye-movements while studying spelling, gives results analogous to the results of similar investigations of the reading process. The adults have, on the average, a broader recognition span, are decidedly briefer in the duration of fixations and make fewer regressive movements.

2. The adults tend to study by syllables but make a detailed analysis of especially difficult parts of a word. Individuals within each of the two groups of sixth-grade students differ in the characteristic unit of study.

3. The adults study principally by repeated surveys of the word as a whole, with regressions occurring primarily at difficult parts. This method of study is also characteristic of much of the work of the quicker learners of the sixth-grade group.

4. The good and poor spellers of the sixth grade apparently differ mainly in the more marked tendency of the good spellers to recognize difficulties and to make a systematic attack in studying words.

5. For the three groups of subjects, but not for the individuals within any one of the groups, there is negative relationship between time spent in study and the amount learned. The adult subjects, with the greatest gain from study, have the shortest learning time; the poor spellers, with the least gain, have the longest.

6. While the results are not conclusive, they suggest the possibilities of this type of study. Analysis of eye-movements probably needs to be supplemented by a more subjective method of analysis. The latter type of analysis might, for example, explain the persistence of errors on the part of poor spellers.

PREDICTING ABILITY IN ART

ALFRED S. LEWERENZ

Statistician, Psychology and Educational Research Division,
Los Angeles City Schools

Purpose

The purpose of this study was to determine with what accuracy ability of high school students in art might be predicted by means of objective tests given prior to taking certain elective courses in art.

Method

At the beginning of school in September 1928, students in a Los Angeles City High school who had just enrolled in classes for Cartooning, Design, Leather, and Crafts were given the Los Angeles Tests in Fundamental Abilities of Visual Art and the results recorded.[1] In February 1929, the semester grades of these students were compared with their performance on the tests five months earlier.

Results

It was found that after the elimination of students who had dropped their course or were in two or more classes, that there were complete data for forty-two individuals. Two sets of figures were available. One set of data was made up of test ratings and semester grades. It was possible for the students to receive one of five ratings on the test based on the normal curve with the following distribution of measures:

Test Rating	Per Cent of Total
1. Very superior	7
2. Superior	24
3. Average	38
4. Inferior	24
5. Very inferior	7
Total	100

The semester grades given in the classes were A, B, C, or D with the following interpretation: A, Excellent; B, Good; C, Fair; D, Passing.

[1] Published by Research Service Co., Los Angeles, Calif.

The second set of data was made up of rankings from 1 to 42 based upon test scores and again upon teacher's estimate.

The first step was to find the average test rating of the pupils receiving grades of A, B, C, and D. Table I gives a summary of test averages by semester grade.

TABLE I

NUMBER OF PUPILS	SEMESTER GRADE	AVERAGE TEST RATINGS
8	A	Very superior
20	B	Superior
13	C	Average
1	D	Superior

The pupil who received D actually had artistic ability according to the teacher but because of behavior difficulties did not make any attempt to use his talent in class.

It will be seen that with the one exception noted average test ratings correspond to semester grades.

The second step was to find the average grade for pupils who received one of the five possible ratings on the test. The results of this grouping are given below.

TABLE II.—TEST RATING

	VERY SUPERIOR	SUPERIOR	AVERAGE	INFERIOR
	4 A	4 A	0 A	0 A
	7 B	9 B	4 B	0 B
	2 C	4 C	5 C	2 C
	0 D	1 D	0 D	0 D
Average semester grade	B +	B −	C +	C

Table II indicates that on the average those who recieved Very Superior ratings on the test also received higher semester grades than those who attained a Superior rating. The same fact is true of the Average and Inferior groups.

Using the second set of data the rankings based upon teacher's estimate were correlated with rankings based upon test scores. It was found that the Pearson Product-moment method gave r as .63($\pm$ .063). This correlation shows that there is a high degree of relationship between scores received on the test and the teacher's estimate made on the basis of achievement five months later.

Conclusions

From the data supplied by the forty-two high school students taking the Los Angeles Tests in Fundamental Art Abilities and who, five months later, received a grade for work done in an elective art course, the following conclusions are drawn:

1. Pupils who did A work received a Very Superior rating on the tests on the average; B students, Superior; and C students, Average.

2. Reversing the comparison, pupils rated on the tests on Very Superior received an average grade of B; Superior, B−; Average, C+; and Inferior, C.

3. A correlation based on test ranking and teacher estimate gave r as .63($\pm$.063) whish shows a high degree of relationship between predicted ability and later performance.

An analysis of marked differences between test score and teacher's rating revealed that in many cases the pupils actually had ability but because of temperamental difficulties had not done their best work.

NOTE ON THE STANDARD ERROR OF A RELIABILITY COEFFICIENT ESTIMATED FROM THE COEFFICIENT FOR A DIFFERENT RANGE OF TALENT

JACK W. DUNLAP AND EDWARD E. CURETON

Territorial Normal and Training School, Honolulu, Hawaii

Kelley[1] gives the formula for estimating the reliability of a test over one range of talent, when its reliability over another range is known, as

$$\frac{\sigma}{\Sigma} = \frac{\sqrt{(1-R)}}{\sqrt{(1-r)}} \tag{1}$$

where σ is the standard deviation for the shorter range

Σ is the standard deviation for the longer range
r is the reliability coefficient for the shorter range
R is the reliability coefficient for the longer range

From (1), we see that

$$R = 1 - \frac{\sigma^2(1-r)}{\Sigma^2} \tag{2}$$

and

$$r = 1 - \frac{\Sigma^2(1-R)}{\sigma^2} \tag{3}$$

The present paper presents the standard errors of formulas (2) and (3), which are the forms in which the formula is actually used.

Let n be the population for the shorter range
N be the population for the longer range

From formula (2), taking differentials,

$$dR = -\frac{2\Sigma^2\sigma(1-r)d\sigma - \Sigma^2\sigma^2 dr - 2\Sigma\sigma^2(1-r)d\Sigma}{\Sigma^4}$$

$$= \frac{2\sigma^2(1-r)d\Sigma}{\Sigma^3} + \frac{\sigma^2\Sigma dr}{\Sigma^3} - \frac{2\sigma\Sigma(1-r)d\sigma}{\Sigma^3}$$

[1] Reliability of Test Scores, *Journal Educational Psychology*, Vol. XI, No. 5, May, 1921, p. 370.

Squaring, summing and dividing by the theoretical infinite population,

$$\sigma_R^2 = \frac{4\sigma^4(1-r)^2\sigma^2_\Sigma}{\Sigma^6} + \frac{\sigma^4\Sigma^2\sigma^2_r}{\Sigma^6} + \frac{4\sigma^2\Sigma^2(1-r)^2\sigma^2_\sigma} + 0 - 0 - \frac{4\sigma^3\Sigma^2(1-r)r_{r\sigma}\sigma_r\sigma_\sigma}{\Sigma^6}$$

The fourth and fifth terms are equal to zero because there is no correlation between variables taken from different populations.

Substituting the values of σ_σ, σ_Σ, σ_r, and $r_{r\sigma}$,

$$\sigma^2_R = \frac{4\sigma^4\Sigma^2(1-r)^2}{2N\Sigma^6} + \frac{\sigma^4\Sigma^2(1-r^2)^2}{n\Sigma^6} + \frac{4\sigma^4\Sigma^2(1-r)^2}{2n\Sigma^6} - \frac{4\sigma^4\Sigma^2r(1-r)(1-r^2)}{2n\Sigma^6}$$

Simplifying,

$$\sigma^2_R = \frac{\sigma^4(1-r)^2}{\Sigma^4}\left[\frac{2}{N} + \frac{3-r^2}{n}\right]$$

$$\sigma_R = \frac{\sigma^2(1-r)}{\Sigma^2}\left[\frac{2}{N} + \frac{3-r^2}{n}\right]^{1/2}$$

Substituting in the first factor the value of r from (3),

$$\sigma_R = [1-R]\left[\frac{2}{N} + \frac{3-r^2}{n}\right]^{1/2} \tag{4}$$

By a similar line of reasoning,

$$\sigma_r = [1-r]\left[\frac{2}{n} + \frac{3-R^2}{N}\right]^{1/2} \tag{5}$$

Formulas (4) and (5) give the standard errors of (2) and (3) respectively.

NOTES ON ARTICLES IN EDUCATIONAL PSYCHOLOGY IN CURRENT ISSUES OF OTHER MAGAZINES

REPORTED BY JAMES E. MENDENHALL

Research Associate, Institute of Educational Research, Teachers College
Columbia University

Intelligence Testing

Measuring Abstract Intelligence in Personnel Work. Public Personnel Studies, May–June, 1929, 69–92. A comprehensive report of types of intelligence, units used, the overlapping of function measured by various tests, lists and discussion of best tests, uses and limitations, weighting, percentile ratings, and the coaching problem.

A Conversion Scale for Comparing Scores on Three Secondary School Intelligence Tests. Robert D. Cole. Journal of Educational Research, October, 1929, 190–198. An absolute scale for comparing scores on Terman, Otis Advanced, and Otis Self-administering tests based upon 6550 scores by boys ages thirteen to nineteen.

Ability Grouping in the First Grade. Hugh J. Bonar. The Elementary School Journal, May, 1929, 703–706. Three groups, (1) upper third (2) one-half of all (3) lower half, were tested for achievement. There was no advantage in segregating either accelerated or slow groups.

Intelligence and Suggestibility in Whites and Negroes. Paul Campbell Young. The Journal of Comparative Psychology, October, 1929, 339–359. A total of 637 children ages nine and ten in the third grade and above were tested with the National Intelligence Test A, two series of lines (one increasing in length and the other decreasing), and several varied forms of a "yes-no" series of answers to questions. In the suggestion tests the Negroes proved much more suggestible than the whites.

Measurement of Achievement

Prognosis Tests in the Modern Foreign Languages. American and Canadian Committee On Modern Languages. The Macmillan Company, New York, 1929. An important report of studies in prognostic tests made by V. A. C. Henmon, Carl C. Brigham, George A. Rice, John W, Todd, John E. Bohan, L. Thomas Hopkins, Percival M. Symonds, and R. J. Van Tassel.

A Silent Reading Test in French. M. Eustace Brown and L. P. Brown. California Quarterly of Secondary Education, October, 1929, 71–76. A report of the content selection, norms, reliability, and related data appears in the article.

A Personnel Survey of 10,000 *Iowa High School Seniors.* J. R. Gerberich and G. D. Stoddard. School and Society, October 12, 1921, 515–520. This survey invòlved testing with a high school content, a comprehension, English aptitude and training, mathematics aptitude, and Thorndike examinations.

Minor Studies on Objective Examination Methods. II. The Theory of Sampling as Applied to Examinations. E. O. Talbott and G. M. Ruch, Journal of Educational Research, October, 1929, 199–206. Experimental studies showed (1) that the essay question called forth less than half (two-fifths) the pupil's knowledge; (2) that the essay examination required twice as much time as the objective examination to treat exhaustively a given unit of subject-matter.

The Oral versus the Mimeographed True-false. Harvey C. Lehman. School and Society, October 5, 1929, 470–472. With classes ranging from twenty-one to twenty-six students in educational psychology, the writer found the oral method slightly more reliable than the mimeographed; and fewer errors by the oral method.

The Effects of Different Directions and Scoring Methods on the Reliability of a True-false Test. Jack W. Dunlap, Adrian de Mello, and Edward E. Cureton. School and Society, September 14, 1929, 378–382. Test means and reliabilities for "guess," "do not guess," and "guess to equalize the number of 'yes' answers and 'no' answers" were calculated.

A-D-U Tests and Examinations. Elizabeth Mason. The Historical Outlook, November 1929, 331–337. A three response type of test with these alternatives, "agreement," "disagreement," and "uncertain" was used. The answer sheet provided three columns of responses: first response, "key," and final response. This type of test proved an excellent learning device.

The Reliability and Validity of Four Types of Vocabulary Tests. Verner Martin Sims. Journal of Educational Research, September, 1929, 91–96. The subjects, one hundred ten school children ranging from Grades V to VIII, were tested with a checking test, the Thorndike multiple-response test, a matching test, and an identification test (Terman vocabulary technique). The reliability of the four types of tests was equally high; the inter-correlation among the tests was high for the identification, multiple-response and matching tests but low for the checking test; the matching test proved most satisfactory.

An Experimental Evaluation of Seven Composition Scales. Dorothy Speer. The Johns Hopkins University Studies in Education, No. 14, 1929. Using such criteria as objectivity of scale, preference of judges, etc., these scales were appraised —Hudelson English Composition, Hudelson Typical Composition Ability, Leonard Judgment of Composition Quality Only, Lewis Measuring Special Types of English Composition, Nassau County Supplement to the Hillegas Scale, Van Wagenen English Composition, and Willing Scale for Measuring Written Composition.

Critical Examination of a Standardized Spelling Test. Guy M. Wilson and A. Rebecca Parsons. Educational Administration and Supervision, October, 1929, 494–498. There was little agreement in the grade placement of words on a standard spelling test and their placement on the Thorndike and Commonwealth lists. The two latter showed that many words on the test were much too difficult for children in the early grades.

Validation of Methods of Testing Spelling. Walter Guiler. Journal of Educational Research, October, 1929, 181–189. The study, involving seven hundred eighty-one students, seventh grade through freshman college, showed that written recall was most effective; oral recall, next most effective; and multiple choice least valid.

EDUCATIONAL PSYCHOLOGY

An Analysis of Studies in Educational Theory. Carter V. Good. Educational Administration and Supervision, October, 1929, 519–548. A discussion and complete bibliography of recent literature regarding character education, curriculum, history of education, methods of teaching, objectives and aims, psychology sociology (heredity), philosophy and ethics.

Retroactive Effect and Degree of Similarity. N. Y. Cheng. Journal of Experimental Psychology, October, 1929, 444–449. Two experiments (*N*, twenty-four to twenty-nine subjects) in the learning of nonsense syllables with various letters common.

The Comparative Retention Values of Maze Habits and of Nonsense Syllables. John A. McGeoch and Arthur W. Melton. Journal of Experimental Psychology, October, 1929, 392–414. Twenty-four college subjects were practised for seven experimental periods, one week apart. In saving in trials, retention of nonsense syllable lists was greater; in time, neither syllables nor mazes was superior; in saving in errors neither was superior.

The Interrelations of Speed, Accuracy, and Difficulty. Laurence L. McLeod. Journal of Experimental Psychology, October, 1929, 431–443. The study measured seventy-five undergraduate college students in their ability to spell six hundred sixty-five words. The conclusions were: (1) relative standing in both speed and accuracy was maintained by spellers at various difficulty levels; (2) the correlation of speed and accuracy was for all words, .346; for easy words, .631, and for medium words, .397.

The Relationship of Subjective Experience to the Performance of Eye Movements. M. D. Vernon. The British Journal of Psychology, October, 1929, 161–172.

Eye-movement Records in the Investigation of Study Habits. W. R. Miles and H. M. Bell. Journal of Experimental Psychology, October, 1929, 450–458. With sixteen college subjects, a high correlation (.70) was secured between perception-time per line and scores on a Thorndike reading comprehension test.

Factors Determining Achievement and Grade Location. J. D. Heilman. Journal of Genetic Psychology, September, 1929, 435–457. An investigation of eight hundred twenty-eight school children (age ten years) in educational age, mental age, life age, school attendance in days, socio-economic status of the home, grade location, date of entering the kindergarten and the first grade. "Our results showed that fifty-seven per cent of the variation in educational age was due to mental age or such hereditary factors as had been measured; about seven per cent of the variation was due to the influences of school training and socio-economic status combined, or such environmental factors as had been measured; and about thirty-six per cent was attributable to such hereditary and environmental factors as had not been accounted for." Other interesting conclusions were revealed by the path-coefficient technique.

The Relation of the Home Factor to Achievement and Intelligence Test Scores. Marlin R. Chauncey. Journal of Educational Research, September, 1929, 88–90. A group of two hundred sixty-three pupils, grades VIII and IX, was measured in Stanford Achievement, Multimental Scale, Chronological Age, and Sims Score Card. Correlations of three orders are presented.

The Language Difficulty of Mexican Children. H. T. Manuel and Carrie E. Wright. Journal of Genetic Psychology, September, 1929, 458–468. Language

tests, an English edition and a Spanish edition, were administered to six hundred sixty-nine high school students and two hundred seven college students enrolled in classes studying the Spanish language.

Time Patterns in Prose. A Study in Prose Rhythms Based upon Voice Records. Helen Griffith, Psychological Monographs, Vol. XXXIX, No. 3, 1929.

Experiments on the Influence of the Amount of Reward on Learning on Young Chickens. G. C. Grindley. The British Journal of Psychology, October, 1929, 173–180. "The amount of the reward has some influence on the learning . . . Under certain conditions a transient kind of learning occurs when no reward is given."

Psychology of Learning and School Subjects

A Study of the Relative Teaching Values of Four Common Practices in Correcting Papers. Francis D. Curtis and Gerald G. Woods. The School Review, October, 1929, 615–623. The method entailing (1) checking of own answers as teacher reads from key (2) followed by a class discussion, proved the most effective teaching device,

Summary of Arithmetic Investigations (1928) *I.* Guy M. Buswell. The Elementary School Journal, May, 1929, 691–698.

Diagnostic and Remedial Work in Arithmetic Fundamentals for Intermediate Grades. May Lazar. Board of Education of the City of New York, Bureau of Reference, Research and Statistics. Publication No. 21, 1928. An excellent report of the kinds of errors made by pupils throughout the grades; and of the methods used for overcoming them.

Improving Computational Ability. Walter Scribner Guiler. The Elementary School Journal, October, 1929, 111–116. A comprehensive testing of a small group of children in the seventh grade was folllowed by an effect remedial program.

Drill in Silent Reading for College Freshman. Robert P. Carroll and Charles C. Jacobs. School and Society, November 9, 1929, 656–658. The drill group showed 2.8 times as much gain in ability to secure ideas from paragraphs than did the control group.

Individual Reading Method Adapted to the Grade Readers for Work Type Reading. Martha Sibley. Educational Administration and Supervision, September, 1929, 441–447. The writer describes a practice-test procedure, and presents objective evidence of its effectiveness.

Studies in the Learning of English Expression. No. III. Vocabulary. Percival M. Symonds and Baldwin Lee. Teachers College Record, October, 1929, 50–58. A study of the vocabulary (a total of 6328 words) of thirty-two compositions in which the rating of words on the Thorndike Wordbook list is compared with the rating of the composition on the Hillegas Scale.

Saving Time in Spelling Instruction. C. P. Archer. Journal of Educational Research, September, 1929, 122–131. The experimental groups consisted of five hundred children in each of the fifth and seventh grades. To each group of 100 pupils training was given in one of the following forms: the base form, the "S" form, the "Ed" form, or the "Ing" form. Positive transfer occurred in equal amounts between base, "s," "ed," and "ing" forms, but was not great between "tion" and other forms. Negative transfer occurred from base to "ing" form in the case of final silent "e," also the reverse.

Improving Handwriting Ability. Walter Scribner Guiler. The Elementary School Journal, September, 1929, 56–62. Diagnosis charts were used to analyze difficulty and prescribe remedial instruction for a class of seventh and eighth grade pupils.

Administration and Supervision

The Principal as a Supervisor. Research Bulletin of the National Education Association, Washington, D. C., November, 1929. These topics are treated: supervision, conferences, teachers' meetings, demonstration teaching, pupil guidance and adjustment.

A Diagnostic Chart for Determining the Supervisory Needs of Teachers of Arithmetic. L. J. Brueckner. The Elementary School Journal, September, 1929, 96–103. The chart includes such major topics as (1) background of technical teaching equipment, (2) knowledge of specific objectives and functions of arithmetic, (3) recognition of the function of general method, etc.

The Administration of the Budget in Secondary Schools. W. C. Reavis and Robert Woellner. The School Review, October, 1929, 589–597. Practices in three hundred sixty-four schools are reported.

Extra-curricular Duties. Earl W. Anderson. Educational Research Bulletin, October 9, 1929, 315–318. Questionnaires submitted to ninety-three high school and seventeen elementary school teachers showed no evidence of reduction of teaching load as extra-curricular duties were assigned.

Subject Combinations in the Programs of Teachers in Small Secondary Schools in New York State. M. G. Nelson. The School Review, June, 1929, 426–432. A study of the programs of teaching of 178 principals, and 1128 teachers.

Items of Information Collected by Departments of Public Instruction of Ten Representative States. William C. Reavis. The Elementary School Journal, May, 1929, 666–673. A tabulation and summary of items.

Review of Educational Legislation. Ward W. Keesecker. Bureau of Education, Department of the Interior, Washington, D. C., 1929.

A Comparative Study of the Effects of Different Types of School-building Ventilation on the Health of Pupils. Russell L. C. Butsch. The Elementary School Journal, (I) September, 1929, 12–26; (II) October, 1929, 123–131. Studying large numbers of pupils, these conclusions were obtained: (1) Absence due to respiratory illness is significantly related to age and grade, the lower age and grade having the greater per cent; (2) a significantly larger per cent of absence was associated with mechanical as contrasted with natural ventilation.

Curriculum

Vitalizing the High School Curriculum. Research Bulletin of the National Education Association, September, 1929. A complete report of recent studies pertaining to the high school fields.

Present Specific Objectives in High School Chemistry. Kimber M. Persing. Journal of Chemical Education, November, 1929, 1958–1978. A study based upon three curriculum investigations, six courses of study, and six textbooks.

The *Vocabulary of American History.* A. S. Barr and C. W. Gifford. Journal of Educational Research, September, 1929, 103–121. The writers sampled the

words used in eight standard high school textbooks. The words were classified by reference to the Thorndike Wordbook. A list of the 1900 most frequently used words is presented in the article.

Illustrative Material for Civics. C. O. Lehman. Educational Research Bulletin, October 9, 1929, 311–315. The amount of space allotted to illustrative materials in ten junior high school textbooks in proportion to the total content ranges from six to twenty-four per cent, with a mean of sixteen per cent. Social problems and economic problems receive the greatest emphasis.

Desirable Literature for Children of Kindergarten Age. Journal of Genetic Psychology, September, 1929, 390–414. An extensive study of nineteen children to determine their interest in stories of various types. A rank order list of stories is made on the basis of terms recalled by children, ratings by experts, etc.

A Social Basis for the Teaching of Elementary English Language. Elizabeth W. Baker. The Elementary School Journal, September, 1929, 27–33. A tabulation of one thousand replies by pupils to the direction "List the things which you do outside of school for which you need to talk or write." For spoken English, "conversation," "business," and "courtesy" were most frequently mentioned; for written English, "letters."

A Score Card for Rating Primary Readers. Cyrus D. Mead. Educational Administration and Supervision, September, 1929, 427–430. The scale covers such main headings as objectives, hygienic features, materials (making for interest), vocabulary, and teacher and pupil helps.

Attitudes, Character and Personality Traits

The Composition of Political Attitudes. Gordon W. Allport. The American Journal of Sociology, September, 1929, 220–238. At the time of the 1928 election for president in the United States, a questionnaire of seven pages was given to three hundred seventy-five students in elementary psychology at Dartmouth. These factors were measured: (1) radical-conservative attitudes; (2) amount of information and misinformation possessed; (3) prejudice; (4) college grades. The returns indicated the following: (1) political radicals are (*a*) less prejudicial, (*b*) have less misinformation, (*c*) omit items of information when uncertain of answers, (*d*) have higher college grades, (*e*) disagree with father's vote, (*f*) are atypical in religious affiliation, (*g*) possess a certain distinctive insight and aggressiveness, (*h*) feel less strongly about the campaign, (*i*) are not altogether consistently liberal, (*j*) consider more important ultimate issues, *e.g.*, government, ownership. These facts are also true of the high-scholarship men and of the low-prejudice group.

Objective Tests, the Best Discoverer of Pupil Attitudes. Samuel Everett. The Historical Outlook, November, 1929, 335–337. A preliminary report of a test to measure the attitudes of high school pupils toward American history.

Group Factors in Temperament Qualities. David W. Oates. The British Journal of Psychology, Oct., 1929, 118–136. The writer tested fifty subjects with a battery of temperament tests, scholastic tests, intelligence tests, and handicraft tests. The intercorrelations indicate the presence of a general intellectual factor and a general emotional factor.

Tests of Temperament and Personality. P. E. Vernon. The British Journal of Psychology, October, 1929, 97–117. A critical discussion of the Downey test is included with a description of the writer's own series of tests.

Personality Differences and Mental Health. William H. Burnham. Journal of Genetic Psychology, September, 1929, 361–389. The writer makes a case for the more extensive use of the genetic method of studying personality.

Moral Judgments of First-grade Children and a Method of Selecting Them. A. M. Carmichael. The Elementary School Journal, May, 1929. The test was given to five hundred four first-grade children. The technique involved a real description of nineteen situations with three responses to each. From the three responses the children selected that which represented the facts. On the several situations, from forty-two to two hundred four (average ninety-six) out of five hundred responses did not represent the facts.

Written Tests of Honesty (Integrity). Public Personnel Studies, July, 1929, 98–106. Tests of change in answers were given three times under different incentives to five groups of testees (custodial officers, police officers, university students, commercial school students, and money handlers).

One Measure of Outcomes of Instruction in the Technique of Teaching. M. E. Herriott. Educational Administration and Supervision, October, 1929, 514–518. Judgments of importance of certain teaching characteristics (adjustment of shades, orderliness of pupils' work, etc.) were secured. The undergraduates' judgments of importance agreed more closely with those of experienced graduate students at the end than at the beginning of the course.

What Characteristics in Teachers Impress Themselves Most Upon Elementary and High School Students? Carl P. Birkelo. Educational Administration and Supervision, September, 1929, 453–456. The following traits were rated most highly: skill in teaching, ability to make points clear in instruction, knowledge of subject-matter, pleasing personality, and ability to inspire student to do his best.

A Study of Personal and Social Traits in Relation to High School Teaching. Floyd Jordan. The Journal of Educational Sociology, September, 1929, 27–43. The concensus of judgment of one hundred fifty high school seniors, one hundred twenty teachers, one hundred supervisors, and one hundred twenty school patrons ranked these traits highest—intelligent, fair, broad-minded, tactful, patient, interesting; and these traits lowest,—modest, proud of profession, reverent, of broad interests, willing to lead, and thought centered outside of self.

Child Psychology

A Preliminary Study of the Factors Affecting the Time Taken by Nursery School Children to Eat Their Food. Flora M. Thurston. Journal of Genetic Psychology, June, 1929, 303–318. A study of forty-one children ages one year and seven months to three years and ten months revealed differences in time due to kind of food eaten, method food was prepared, and age differences.

Child Development Abstracts and Bibliography. Committee on Child Development, National Research Council, Washington, C. D., October, 1929.

Some Phases of Nursery-kindergarten-primary Education, 1926–1928. Mary Dabney Davis. Bulletin No. 29, Bureau of Education, Department of the Interior, Washington, D. C., 1929.

Exceptional Children

A Brief Survey of Special Education in the Public Schools of Baltimore. J. E. Wallin. Department of Education, Baltimore, 1929. A comprehensive survey of the present provision and needs in special education.

A Study of Juvenile Theft. Hary J. Baker, Fred J. Decker, and Arthur S. Hill. Journal of Educational Research, September, 1929, 81–87. An experimental group of cases (*N*, eighty-four boys) formally convicted of theft by a juvenile court and a control group of equivalent size, age, etc. displayed the following: (1) no significant differences in economic status, father's occupation, size of family, position of boy among siblings, roomers or boarders in the home church attendance, health or injury, time in school, change of schools, school marks, work and earning power of boys, and attendance at movies; (2) differences in favor of the control group were conventional ages of parents at the birth of their children, unbroken homes, lack of crowding in homes, general intelligence, supervision of play, correction of physical defects, suitable playmates, and church affiliation.

The Behavior-problem Boy. Albert A. Owens. Journal of Educational Research, October, 1929, 166–180. A study of three hundred sixty-five boys sent to a disciplinary school showed that the leading causes of transfer from regular schools were the following offenses: truancy, forty-nine per cent of the cases; disobedience, forty-five per cent; fighting or bullying, twenty-five per cent; disorder, twenty-four per cent; defiance, nineteen per cent; impudence, nineteen per cent. About fifty per cent of the group has a court record; physically the group was normal; the average IQ was between 75; and 80.

The Need for Guidance in Orphan Homes. Robert A. Davis. The Elementary School Journal, September, 1929, 51–55. The study included a mental survey with the Dearborn Group (I and II), and the Haggerty (Delta I and II). IQ distributions are presented. A check chart shows the types of occupational activities provided by different orphanages.

Growth in Height of Feeble-minded Children. C. Wilson Smith. Journal of Genetic Psychology, June, 1929, 330–341. A report of a six-year study of normal and of feeble-minded children shows that sub-normal children grow more slowly than do normal children, and are distinctly shorter as a group than normal children.

Educational and Vocational Guidance

A Study of Vocational Interests of California High School Students Based on a Survey of Twelve Rural High Schools. J. H. Bedford. California Quarterly of Secondary Education, October, 1929, 47–66. The results of this study show the striking discrepancy between vocational interests of students and the occupational needs of society.

Guidance Practice in Junior Colleges. Frederick J. Weersing and Leonard V. Koos. California Quarterly of Secondary Education, October, 1929, 93–104.

Trends in College Entrance Offerings. F. A. Balyeat. The School Review, June, 1929, 451–456. Trends from 1907 to 1917 to 1927 in the University of Oklahoma are included in the article.

The Problem of Elimination from College. Noel B. Cuff. School and Society, October 19, 1929, 550–552. A study of two hundred ninety-two freshmen showed various factors associated with elimination.

The Relation of Success in Certain Subjects in High School to Success to the Same Subjects in College. Royal Gilkey. The School Review, October, 1929, 576–588. Correlations (*N*, 90 to 210) between high school marks in certain subjects and college marks in the same subjects ranged from .151 to .498 with an average of .35.

Some Reasons Why High School Pupils Choose Certain Subjects. Silas E. Gould and Robert A. Davis. School Review, October, 1929, 602–614. The most important reasons were in order: (1) as an aid in making a living; (2) because of advice from parents, friends, or schoolmates; (3) because I believe I have the ability to make a good mark in the subject; (4) because of advice from principal, teacher, or superintendent; and (5) because of the reputation or standing of the subject.

The Scholarship of University Students Participating in Extra-curricular Activities. James E. Knox and Robert A. Davis. Educational Administration and Supervision, October, 1929, 481–493. A study of 5377 college students showed that the active students were superior in scholarship to non-active. The order of superiority within the active groups was (1) publication, 81 points; (2) administration and executive, 81 points; (3) college representation, 79.4 points, and (4) athletic, 77.6 points.

A Study of Certain Phases of the Conduct of Student News Publications. M. J. Nelson. School and Society, November 2, 1929, 609–610. The returns from a questionnaire submitted to fifty-eight colleges and universities.

Greeks of the Campus. Clifford L. Constance. School and Society, September 21, 1929, 409–414. "Fraternity freshmen at the University of Oregon are about the same as non-fraternity freshmen on the basis of the psychological test score, are inferior to them in grades earned in high school, but on the average secure better grades in the university. Sororities show the same behavior, except that their freshmen are better than the non-sorority freshmen women on the mental test basis."

Special Abilities

A New Method of Studying the Rhythmic Responses of Children Together with an Evaluation of the Method of Simple Observation. Christian Paul Heinlein. Journal of Genetic Psychology, June, 1929, 205–228. An interesting study employing a player piano, a special platform, and a kymograph to determine the rhythmic responses of young children in walking.

A Survey of Studies in Experimental Aesthetics. R. Alice Drought. Journal of Educational Research, September, 1929, 97–102. The study included the judgments of one hundred sixty adults and sixty-six fifth grade pupils regarding a battery of nine sets of pictures, five to a set. One of the five pictures was termed the "Perfect" copy; each of the other four violated some one principle of art—unity, harmony, proportion, or congruity. The groups seemed less annoyed by a violation of unity than by a violation of harmony or of proportion.

Miscellaneous

Creative Education and the Science of Education in America. Frank N. Freeman. Chicago Schools Journal, September, 1929, 1–6. A critical article emphasizing the need for both the creative and the scientific in education.

The Relation of Philosophy and Science in Education. H. Gordon Hullfish, Journal of Educational Research, October, 1929, 159–165. "Science and philosophy are but different ways of dealing with facts. The former handles its facts, to discover *what is;* the latter with reference to the same facts, wants to determine *what should be.*"

Some Facts about College Presidents. Homer P. Rainey. School and Society, October 26, 1929, 580–584. A study of previous vocations of presidents, of teaching experience, of various subjects taught, of degrees held, of present ages, of length of time in present positions, etc.

Rural Education in 1926–1928. Katherine M. Cook. Bureau of Education, Department of the Interior, Bulletin No. 18, 1929. A summary of tendencies during a two-year period.

The Problem of the Feeble-minded in New Jersey. Research Department, The Training School at Vineland, New Jersey, July, 1928. Three case studies of typical degenerate families are presented.

Sones-Harry High School Achievement Test. World Book Company, Yonkers N. Y., 1929.

NEW PUBLICATIONS IN EDUCATIONAL PSYCHOLOGY AND RELATED FIELDS OF EDUCATION

CONDUCTED BY FRANCES M. FOSTER

The Process of Human Behavior, by Mandell Sherman and Irene Case Sherman. New York: W. W. Norton and Co., 1929. Pp. 227.

"The Process of Human Behavior" is an interesting and useful volume. In scope it is not as wide as its title suggests. It is not a philosophical treatise profoundly concerned with man's destiny. It is not, as one might gather from the content of the publisher's advertising, an attempt by two serious students of the subject to popularize a whole field of knowledge they know at first hand. It is not an attempt to present an exhaustive study for either popular or scholarly purposes. Its aim as the authors give it in the introduction is "to outline the principles underlying the process of human behavior and the important factors which influence it; to describe the working of the human organism as a whole and the genesis and development of the methods by which adjustments of the individual to his environment are affected." Do the authors by their selection of subject matter and presentation of it satisfy this aim? An answer to this question calls for at least a brief analysis of the nature of the content, as well as the organization and presentation of it.

The content of the book is readily classifiable into four large divisions, each presented in two chapters. In the middle four chapters appear the experimental work of one or both of the authors: their researches on sensori-motor development in Chapters III and IV and that of differentiation of emotional responses in infants in Chapters V and VI. Most of the fifteen illustrations appear in the two last named chapters. These researches are prefaced by two chapters on nervous systems, purposed to serve as background materials. The last two chapters are concerned with the nature of personality with emphasis on the clinical approaches to the study of behavior problems. The book, in short, then is composed of the researches of one or both of the authors, on sensorimotor development and differentiation of emotional responses in infants prefaced

by two chapters on neurology and followed by two chapters on personality.

Does it satisfy the aim as given by the authors? No, from the point of view of extensiveness of knowledge necessary for an intelligent understanding of our mental adjustments. Two chapters on neurology in a book intended for popular consumption can hardly be said to represent the trend of the course of events in psychology. The presentation of the lowering resistance as the neurological explanation for learning is not exactly in accord with Dr. Lashley's experimental findings. However, the emphasis in these chapters is on phylogeny as a means of throwing light on the nature of the increasingly adaptive behavior we find as we go up the scale of the animal series . . . In their chapters on sensorimotor development in spite of the fact that the second of the two chapters on the subject is called "The Relation of Sensori-motor Development to the Growth of Intelligence" nothing is said of the work of other investigators, not even Gesell. In their discussion of language development we find not a word about the work of Piaget, not to mention others. In their chapters on emotions we again find the work of others not mentioned. Moreover, the cultural anthropologist can justifiably say of this book that it does not begin to throw light on the nature of actual behavior of adults. The subject-matter, then, is very limited in scope but that little, we hasten to add, is well presented and all of it serves as an object lesson in the objective and biological approach the authors strongly advocate in their introduction. H. MELTZER.

Psychiatric Clinic, Saint Louis, Missouri.

The Child in Primitive Society, by Nathan Miller. New York: Brentanos, 1929.

In "The Child in Primitive Society" the author seeks to assemble, consolidate, and in some degree to interpret a tremendous mass of material relating to the training of children in societies at an early stage of social development. It is a scholarly and comprehensive piece of work. The numberless examples given and the fat sections devoted to references and bibliography show the painstaking exhaustiveness which characterizes the best in anthropological research. This scholarship does not, however, weigh heavily upon the neurones of the reader. The book is interesting, coherent and is likely to be read by many who do not "have to."

The examination of education in primitive societies leaves the educational reader with a firmer conviction that progress in education is a rather sophisticated business, and is not to be achieved by any return to conditions of imagined primitive "naturalness." The primitive child receives a wealth of misinformation, bad habits and inhibitions exceeding that of a child reared in the most conventional of modern institutions. There is plenty of naïveness but really little naturalness in it. While the study of primitive societies tells much concerning the origin of customs and institutions, the study of the present child by exact methods gives far sounder information concerning his true nature.

This is one of the first publications of a "Library of Educational Psychology" sponsored by Kegan Paul in London and by Brentano's in this country. Other volumes in preparation give promise that the series may compare favorably in its field with the "International Library of Psychology, Philosophy and Scientific Method" which has performed such valuable service to contributers and to readers in the wider aspects of psychology.

LAURANCE F. SHAFFER.
Carnegie Institute of Technology.

Teaching in College and University; A Survey of the Problems and Literature in Higher Education, by Carter V. Good. Baltimore: Warwick and York, Inc., 1929. Pp. XX + 557.

The teacher-training movement began in the elementary school and has in recent years spread into the higher levels of the educational system. The most recent step in this movement has been the trend toward better professional training of college teachers. The common statement that we find the best teaching in the elementary school, next best in high school, and the poorest in the college and university, whether strictly true or not, shows that there is a growing tendency to expect college teachers to improve their technique of instruction.

Dr. Good's new book is a very fine contribution to this movement toward the professional training of college teachers. It brings together into one volume the essence of what has been written about higher educational problems. This literature, scattered through numerous volumes and countless magazines, has heretofore been difficult of access, and should now take on new usefulness to educational workers.

The book is organized into eight chapters, as follows: I. Evidences of Interest in College Teaching: Sources of Information; II. Status of

Colleges, Students, and Teachers; III. Objectives and Standards in Higher Education; IV. Selection and Organization of Subject-matter: The Curriculum; V. Psychology of Learning: Types and Economical Bases of Learning; VI. Conduct of the Class Period, Study Problems, and Provision for Individual Differences; VII. Measurement and Guidance in Higher Education; VIII. Reorganization and Adjustment in Higher Education, Improvement of College Teaching, and Research. In addition to these chapters the book contains an appendix of eighty-five pages, consisting of bibliographies on special methods and professional problems in the various subject fields.

The style of organization and writing is comparable to that in Dr. Good's recent work "How to Do Research in Education." There are seven hundred ninety numbered footnotes in the book, many of which contain as many as ten or fifteen different bibliographical citations. In addition, there is a long list of supplementary references at the end of each chapter. The author does not say much about college problems himself. He devotes almost all of his space to the presentation of the views and contributions of the writers who have dealt with topics in this field. The book might therefore be classed as a splendid piece of bibliographical research rather than as a textbook of the conventional type.

It will prove invaluable to scholars who are making studies of college and university problems, as well as to the students and instructors in the growing number of courses in methods of college teaching.

The book has one limitation, as seen by the reviewer, and that is not the fault of the writer. It distributes the emphasis on the different topics in very much the same way as it is distributed in the articles and books that are summarized. The result is that much space is devoted to certain administrative problems of interest to presidents and deans but not to professors and classroom instructors. Many of the latter will wish that more articles and books might be written on *how to teach in college* instead of on *how to operate a college.* Probably when Dr. Good's book is revised in the light of the literature that is to be published in the next decade we shall see a larger amount of space given to problems of actual classroom instruction. Dr. Good has presented the literature as it is now; let us hope that writers on college education in the future will devote more attention to the instructional side of the problem.

C. C. Crawford.

University of Southern California.

INCHES
1
2
3
4
5
6
C2399
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
METRIC